Theses and Dissertations   1. Thesis and Dissertation Collection, all items

1995-09

NPSNET-3D sound server: an effective use of the auditory channel

Storms, Russell L.

Monterey, California. Naval Postgraduate School

http://hdl.handle.net/10945/35203

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

NPSNET-3D SOUND SERVER:
AN EFFECTIVE USE OF THE AUDITORY CHANNEL

by

Russell L. Storms

September 1995

Thesis Advisor:	Michael J. Zyda
Thesis Co-Advisor:	John S. Falby

Approved for public release; distribution is unlimited.

19960226 130   DTIC QUALITY INSPECTED 1

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time reviewing instructions, searching existing data sources gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE September 1995	3. REPORT TYPE AND DATES COVERED Master's Thesis

4. TITLE AND SUBTITLE
NPSNET-3D SOUND SERVER: AN EFFECTIVE USE OF THE AUDITORY CHANNEL(U)

5. FUNDING NUMBERS

6. AUTHOR(S)
Storms, Russell L.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/ MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/ MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the United States Government.

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The current MIDI-based sound system for the distributed virtual environment of NPSNET can only generate aural cues via free-field format in two dimensions. To increase the effectiveness of the auditory channel in NPSNET, a sound system is needed which can generate aural cues via free-field format in three dimensions.

The approach taken was to build upon the current NPSNET sound system: NPSNET-PAS [ROES94]. Hardware limitations of NPSNET-PAS sound generating equipment were identified and more capable "off-the-shelf" sound equipment was procured. In software, a new algorithm was developed which properly distributes the total volume of a virtual sound source to a cube-like configuration of eight loudspeakers. A second algorithm, based on the "Precedence Effect," was also developed in an attempt to enhance one's ability to localize a sound source. Synthetic reverberation using digital signal processors was added to enhance perceptual distance of the generated aural cues.

The result of this research is a MIDI-based free-field sound system consisting of "off-the-shelf" sound equipment and computer software capable of generating aural cues in three dimensions for use in NPSNET. This sound system was tested during numerous demonstrations of NPSNET and proved capable of generating eight independent audio channels required for potential output to a cube-like configuration of eight loudspeakers laying the foundation for increasing one's level of immersion in NPSNET.

14. SUBJECT TERMS
Sound Localization, Spatial Audio, 3D Sound, Psychoacoustics, HRTF, Binaural Recordings, MIDI, Free-Field, NPSNET, DIS, Virtual Environment, Auditory Channel, Virtual Audio.

15. NUMBER OF PAGES
191

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

NPSNET-3D SOUND SERVER:
AN EFFECTIVE USE OF THE AUDITORY CHANNEL

Russell L. Storms
Captain, United States Army
B.S., United States Military Academy, 1986

Submitted in partial fulfillment of the
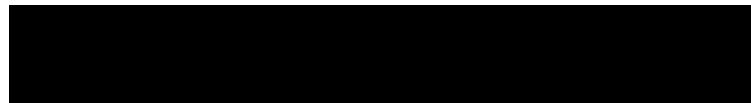requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL

September 1995

Author:

Russell L. Storms

Approved By:

Michael J. Zyda, Thesis Advisor

John S. Falby, Thesis Co-Advisor

Ted Lewis, Chairman,
Department of Computer Science

ABSTRACT

The current MIDI-based sound system for the distributed virtual environment of NPSNET can only generate aural cues via free-field format in two dimensions. To increase the effectiveness of the auditory channel in NPSNET, a sound system is needed which can generate aural cues via free-field format in three dimensions.

The approach taken was to build upon the current NPSNET sound system: NPSNET-PAS [ROES94]. Hardware limitations of NPSNET-PAS sound generating equipment were identified and more capable "off-the-shelf" sound equipment was procured. In software, a new algorithm was developed which properly distributes the total volume of a virtual sound source to a cube-like configuration of eight loudspeakers. A second algorithm, based on the "Precedence Effect," was also developed in an attempt to enhance one's ability to localize a sound source. Synthetic reverberation using digital signal processors was added to enhance perceptual distance of the generated aural cues.

The result of this research is a MIDI-based free-field sound system consisting of "off-the-shelf" sound equipment and computer software capable of generating aural cues in three dimensions for use in NPSNET. This sound system was tested during numerous demonstrations of NPSNET and proved capable of generating eight independent audio channels required for potential output to a cube-like configuration of eight loudspeakers laying the foundation for increasing one's level of immersion in NPSNET.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

At first thought, I did not think that it was important to list any acknowledgments. Because typically, people seem to just quickly scan over the acknowledgments and get to the meat of the subject. I am one of these people. However, it occurred to me that the acknowledgments are not really for the benefit of those actually listed in the acknowledgments, but rather for the author -- me. The reality is that in ten years no one will probably be reading this thesis except perhaps me. As a result, the acknowledgments will not only help to remind me of those who had helped me with this thesis, but it will also help to remind me of my thoughts pertaining to this thesis and of life itself. So here are my acknowledgments, if not for posterity's sake, then for my own.

First of all, this thesis would be have been possible in the first place, if John Roesli had not taken the time to tell me what he was doing playing the EMAX II keyboard in the early morning hours in the Graphics and Video Laboratory. Thanks to the folks at E-mu, Ensoniq, Lexicon, and Crystal River Engineering for answering my numerous technical questions. A special thanks goes to Brain Trankle of Trankle Associates for his advice in configuring sound equipment. Also, thanks to Chris Gunn of *On The One Productions* for his professional production recording advice. Thanks to David Burgess for inviting me to Interval Research to exchange ideas on virtual audio. I would also like to personally thank Dr. Elizabeth Wenzel and Dr. Durand Begault of NASA-Ames Research Center for lending me their ears every now and then. Thanks for the great insights that I learned from the folks at CCRMA especially Brent Gillespie, Craig Sapp and Sila O'Modhrain. A very special thanks to Henry Ong for deriving the mathematics behind the sound cube model. Thanks to Lloyd Biggs for his support with my never ending configuration changes of the sound system. Thanks to John Falby for his meticulous attention to grammatical details. Also, thanks to Paul Barham who had an answer to everyone of my programming questions about NPSNET and even some that I did not ask for. However, even after all the thanks I have given thus far, the overwhelming majority of my thanks goes to my thesis advisor Dr.

Michael Zyda. His faith and support in me are what made this thesis possible. Whenever I needed anything, be it hardware, software, advice, or just someone to talk to, Dr. Zyda always came through with more than I expected. Thanks for everything Dr. Zyda.

At this point, I would like to change my viewpoint. I apologize if I have forgotten to give thanks to someone, but I must move on to the most important factor in my accomplishing this thesis -- my wife. Thank you Deanna for dealing with my very long hours away from home in the computer lab. However, you could have complained a little less. The ironic thing is that your love was always a constant reminder to me that things like computer projects, exams, and even this thesis mean absolutely nothing in the big picture of life. Your love is the most important thing in my life, and no computer project, exam, or thesis will ever compare. I have cherished every moment together and I cannot wait to spend everyday of the rest of my life with you and our soon to be born child, Janell Renae.

I. INTRODUCTION

The primary objective of much research over the years in the virtual reality community has been to improve three-dimensional (3D) visual simulation cues. However, to augment one's immersion in a virtual environment, audio cues are a vital complement. To be most effective, these audio cues should be presented in 3D as opposed to 2D. These 3D audio cues are commonly known as spatialized audio or 3D sound and represent a rapidly growing area of interest in the field of virtual reality [DURL95]. This growing interest has produced numerous theories and working applications of 3D sound systems for use in various virtual environments.

A.   MOTIVATION

The primary motivation of this thesis was to design and implement an appropriate 3D sound system for use with the Naval Postgraduate School Networked Vehicle Simulator (NPSNET) [ZYDA93] [ZYDA94] [MACE94]. NPSNET is an ongoing research effort by the NPSNET Research Group (NRG) conducted with the resources of the Graphics and Video Laboratory in the Department of Computer Science at the Naval Postgraduate School (NPS) in Monterey, California. NPSNET is the first 3D virtual environment suitable for multi-player participation over the Internet. It uses IP multicast network protocols and the IEEE 1278 Distributed Interactive Simulation (DIS) application protocol [DEER89] [IEEE93]. NPSNET uses relatively low-cost Silicon Graphics IRIS workstations to produce quality images at the high frame rates required for real-time visual displays. In an effort to keep costs low, a correspondingly low-cost 3D sound system, capable of generating effective real-time 3D audio displays, is needed.

B.   RESEARCH OBJECTIVES

Since 1991, the NRG has developed various theories and working applications for integrating aural cues into the virtual environment of NPSNET [DAHL92] [ROES94].

1

These systems, though very capable, could only generate aural cues in two dimensions. The primary objective of this research is to design and develop a free-field sound system for integrating aural cues in three dimensions into the virtual environment of NPSNET. The resulting sound system is NPSNET-3DSS: Naval Postgraduate School Networked Vehicle Simulator-3D Sound Server.

The previous NPSNET sound system: NPSNET-Polyphonic Audio Spatializer (NPSNET-PAS) was used as the foundation for developing NPSNET-3DSS. In the development of NPSNET-3DSS, a three phase approach was utilized. Phase one considered using only the existing sound equipment previously available in the Graphics and Video Laboratory. The second phase considered using not only the existing sound equipment in the lab, but also considered using a *wish list* of sound equipment that could be purchased in the future. In this phase, extensive research was conducted in order to find sound equipment for a relatively low cost which would enhance, yet still complement, the existing sound system. The third phase, and most difficult, was a combination of the first two phases. This phase considered the realistic possibility that only some of the sound equipment on the *wish list* would be purchased. The difficulty of this approach was not knowing which sound equipment will eventually be available for implementation. Thus, a larger number of possible sound equipment configurations were considered during the theoretical and design phase of this thesis. However, as new sound equipment was eventually purchased from the *wish list*, the number of these possible configurations was reduced.

The following are the preliminary objectives that encompass all three phases.

- Compare and contrast headphone and free-field sound delivery systems.
- Identify current sound equipment limitations and procure better capable sound equipment.
- Design and implement a general mathematical sound model for properly distributing the volume of a virtual sound source to the various loudspeakers in both a 2D and 3D free-field sound system.
- Verify the effectiveness of volume distribution and localization of the new general mathematical sound model through demonstrations of NPSNET.

- Design and implement a sound model based on the *Precedence Effect* for improving the ability to localize a virtual sound source via free-field delivery.

- Evaluate the effectiveness of using binaural recordings presented in free-field format.

- Provide an appropriate direction for future NPSNET sound systems.

- Provide more realistic and better sampled sounds for NPSNET by recording actual sounds in the field at measured distances by means of portable Digital Audio Tape (DAT) recorder.

- Investigate the possibility of moving all generated sounds to one platform, the IRIS Workstation, in order to increase standardization and portability.

C.   SCOPE

The focus of this research is on the theory, development, and practical application of applying aural cues for use within the distributed virtual environment of NPSNET. This research is centered primarily around the question of how to increase one's level of immersion into the virtual world of NPSNET through the use of the auditory channel. To answer this question, relevant software and hardware issues are discussed as they pertain to the design and implementation of a sound system using the Musical Instrument Digital Interface (MIDI) protocol. Furthermore, this research focuses on using commercial *off-the-shelf* sound equipment as opposed to custom designed equipment made specifically for this research effort. The reason for using *off-the-shelf* sound equipment is as follows: 1) for reduced cost; 2) for investigating how commercial market sound equipment can be used to enhance the auditory channel of virtual environments; 3) to ease standardization and portability of this research; and 4) to make the results of this research effort more easily available to those interested. Lastly, it should be noted that this thesis does not focus on such low level areas as digital signal processing design and Fourier analysis. Such low level concepts are indeed relevant in the area of this research and numerous other applications of 3D sound, but are beyond the scope of this research.

D. LIMITATIONS

1. Anechoic Chamber

Since this research centers on the delivery of sound through free-field format, use of an anechoic chamber would greatly improve the ability to measure the effectiveness of the generated auditory displays. Although highly desirable, an anechoic chamber was not available for this research. As a result, the only feasible and practical location for conducting this research was in the Department of Computer Science's Graphics and Video Laboratory located on the fifth floor of Spanagel Hall at the Naval Postgraduate School. This laboratory is typical of most computer labs. It is designed primarily for the purpose of allowing people to use computer workstations. Thus, this research inherently suffers from the poor room acoustics typically associated with computer labs.

2. Common Ground

Another problem with conducting research in the Graphics and Video Laboratory was the lack of a common ground for all electrical devices. As a result, a slight audible hum was intermittently present when operating sound equipment in the lab. Although the presence of this hum would be totally unacceptable in any type of sound generating facility, it did not affect research efforts. It's only affect was degrading the overall quality of generated sound.

3. Lack of Continuity

The Department of Computer Science's Graphics and Video Laboratory does not have a full-time audio lab technician. The only technical audio support provided to the lab has been intermittent part-time audio technicians. Thus, there is a lack of continuity in audio expertise in the lab. As a result, much time was spent inventorying the audio hardware and software that was actually available in the lab and then learning their capabilities and usage.

E.   ASSUMPTIONS

There is no certain level of knowledge that the reader is assumed to possess in order to read and understand this thesis. Practically all the concepts discussed in this research are presented with the layman in mind. However, this research is better understood if the reader has a basic knowledge of computers, virtual worlds, MIDI, audio systems, and acoustics.

F.   LITERATURE REVIEW

In the preparation of this research, a thorough literature review was performed. The results of this review were instrumental in preparing this research and are presented as an annotated list of references which can be found in the bibliography. This list is a conglomeration of references which were gathered from various research efforts including: 1) Elizabeth Wenzel from NASA-Ames Research Center; 2) Richard Duda from San Jose State University; 3) Center for Computer Research in Music and Acoustics (CCRMA) from Stanford University; and 4) the NRG 3D Sound Library at the Naval Postgraduate School. This consolidated list is quite exhaustive including numerous facets of sound as it pertains to various theories and applications. This list is a vital resource for anyone interested in pursuing further research of sound not only as it pertains to its use in virtual environments, but also in practically any application.

G.   THESIS ORGANIZATION

This thesis is organized around twelve chapters and eight appendices. Chapter II outlines the previous work in applying aural cues for use in NPSNET. This chapter is important for it is the first   attempt to document the history of the NPSNET sound servers. The knowledge gained from this chapter helps to understand this current research effort. Chapter III provides a background of the wave properties of sound, 3D sound perception, the decibel, Inverse-Square Law, and MIDI. It is essential for the layman to read and understand this chapter before reading any other chapters. Chapter IV explains the concept of the *auditory channel* and tries to clear-up some of the confusion associated with the

terminology of 3D sound. Chapter V analyzes the advantages and disadvantages of headphones and free-field systems in the application of improving the level of immersion in VEs. Chapter VI gives an overview of the NPSNET-3DSS. Chapter VII gives the derivation of the 3D sound cube model (SCM). Chapter VIII discusses the development of the Precedence Effect (PE) sound model. Chapter IX gives a background and history of the use of synthetic reverberation (SR), and then discusses how SR can be used in VEs to increase distance perception of sound events. Chapter X describes the software and hardware functionality of NPSNET-3DSS. Chapter XI gives the implementation and analysis of the 3D SCM, PE sound model, and SR for use in NPSNET-3DSS. Chapter XII is the concluding chapter which discusses the overall results of this research effort, follow-on work, recommendations, and some final thoughts.

Appendix A contains a list of definitions and abbreviations used throughout this thesis. Appendix B contains the user guide for setting up and running NPSNET-3DSS. Appendix C lists all the hardware wiring diagrams of equipment utilized in this research effort. Appendix D describes how to configure and use the EMAX II for use with NPSNET-3DSS. Appendix E describes the Allen & Heath GL2 mixing board and also how to configure the mixing board for use with NPSNET-3DSS. Appendix F contains information on how to configure the Ensoniq DP/4 to respond to MIDI commands for use in NPSNET-3DSS. Appendix G presents a brief description on binaural recordings. Appendix H describes some experiments on sound perception that were performed at the 1995 CCRMA Summer Workshop: *Introduction to Psychoacoustics and Psychophysics with emphasis on the audio and haptic components of virtual reality design* at Stanford University.

H.   DEFINITIONS AND ABBREVIATIONS

See APPENDIX A: LIST OF DEFFINITIONS AND ABBREVIATIONS on page 119 for a list of definitions and abbreviations relating to pertinent aspects of this research.

II. PREVIOUS WORK

Since 1991, the NPSNET Research Group (NRG) has developed various theories and working applications for integrating aural cues into the virtual environment of NPSNET. Although there are two types of sound delivery systems for which these cues can be generated, headphone systems and free-field systems, all of these previous working applications have presented aural cues via free-field format (i.e. loudspeakers). The advantages and disadvantages of these two types of sound delivery systems are discussed in Chapter V. HEADPHONES VS. FREE-FIELD DELIVERY SYSTEMS. Prior to this research there have been a total of three working sound systems for generating aural cues into NPSNET: 1) NPS-Sound, 2) NPSNET Sound Server, and 3) NPSNET-Polyphonic Audio Spatializer. A common factor in each of these sound systems is the IRIS Workstation by Silicon Graphics Inc. (SGI). Since NPSNET is run on IRIS Workstations, each sound system must have the capability to interface with these SGI machines in real-time. The following is a brief description of these previous sound systems.

A. SOFTWARE TESTBED

Before discussing the details of the previous work in this research area, a little needs to be said about the software testbed. The primary software testbed utilized for all previous and current NPSNET sound systems has been NPSNET. The latest version of this software is NPSNET-IV [ZYD93] [ZYD94] [MAC94]. NPSNET-IV is the first 3D virtual environment suitable for multi-player participation over the Internet. NPSNET-IV uses Internet Protocol (IP) multicast network protocols and the IEEE 1278 Distributed Interactive Simulation (DIS) application protocol [DEE89] [IEE93]. NPSNET is an ongoing research effort by the NRG and has devoted itself to exploring several areas of interactive simulation including [MAC94]:

- Application and network level communication protocols.
- Object-oriented techniques for virtual environment construction.

- Hardware and operating system optimization.

- Real-time physically-based modeling (e.g. smoke, dynamic terrain, and weather).

- Multimedia (audio, video and imagery).

- Artificial intelligence for autonomous agents or entities.

- Integrating robots into virtual worlds.

- Human interface design (e.g. stereo vision and system controls).

NPSNET-IV is unique in distributed simulation. It functions as a fully operational visual simulator providing a research testbed for the above areas while incorporating the following [MAC94]:

- Distributed Interactive Simulation (DIS 2.04) protocol for application level communication among independently developed simulators (e.g. legacy aircraft simulators, constructive models, and real field instrumented vehicles).

- IP Multicast, the Internet standard for network group communication, to support large scale distributed simulation over inter-networks.

- Heterogeneous Parallelism for system level pipelines (e.g. draw, cull, application, and network) and for the development of a high performance network software interface.

B.   NPS-SOUND

The first attempt to add aural cues to NPSNET for the purpose of increasing the listener's level of immersion was in 1991. This first effort was conducted by Joseph Bonsignore, Jr. and Elizabeth McGinn both of whom were Master of Computer Science students at NPS. Because there is no concise documentation of this research effort, the following will be the first attempt to formally document this important research endeavor.

1. Hardware Systems

NPS-Sound consisted of the following equipment:

- One *Macintosh* (MAC) *IIci* computer having a 32-bit Motorola 68030 microprocessor running at 25 MHz with 8 megabytes of RAM.

- One *Quantum* 210 Megabyte external hard drive.

- Two *Syquest* 44 Megabyte removable hard drives.

• Two *Farallon MacRecorders*. These are relatively inexpensive audio digitizers each with a built-in microphone that plugs in to one of the MAC's serial ports. In 1991, a *MacRecorder* with its accompanying software *SoundEdit* cost $249.00. [FARA90] [LEHR91].

• Digidesign's *Sound Designer II*. This is an extensive Macintosh-oriented sound production lab complete with sophisticated sound editing/sound synthesis capabilities. *Sound Designer II* dramatically extends the editing capability of the *MacRecorder*. It includes a DSP chip with sampling rates up to 44.1 KHz (CD quality), an Analog-to-Digital (AD) converter, and its accompanying software *SoundTools*. This is indeed a very powerful system which in 1991 cost $3285.00. [DIGI90] [LEHR91].

• Carver Power Amplifier TFM-6C with 240 watts total power.

• One set (a total of 2) of Infinity Reference Three Speakers.

2. Software

• Opcode's *Studio Vision*. This is also a powerful program which runs on the MAC providing digital-audio recording, editing, and playback. The cost in 1991 was $995.00. [OPCO90] [LEHR91].

• *FontesTalk II*. A *Prograph* program.

• *SoundMover*.

• *Practica Musica*.

• *ConcertWare++*.

3. General Description

The interface between NPSNET and this sound system was an IRIS 4D/240 VGX workstation having four 25 MHz processors and 64 MB of RAM. Based upon certain events, a *C* program which resided on the VGX workstation generated commands as a string to the MAC via an RS-232 serial interface. This string contained the name of an audio file which resided on the MAC. The *Prograph* program, *FontesTalk II*, deciphered the string and played the appropriate audio file. This audio file's signal was sent from the MAC to a Carver power amplifier which was routed to two Infinity speakers ultimately providing the appropriate aural cues to the NPSNET user. See Figure 1 for an overview of this system.
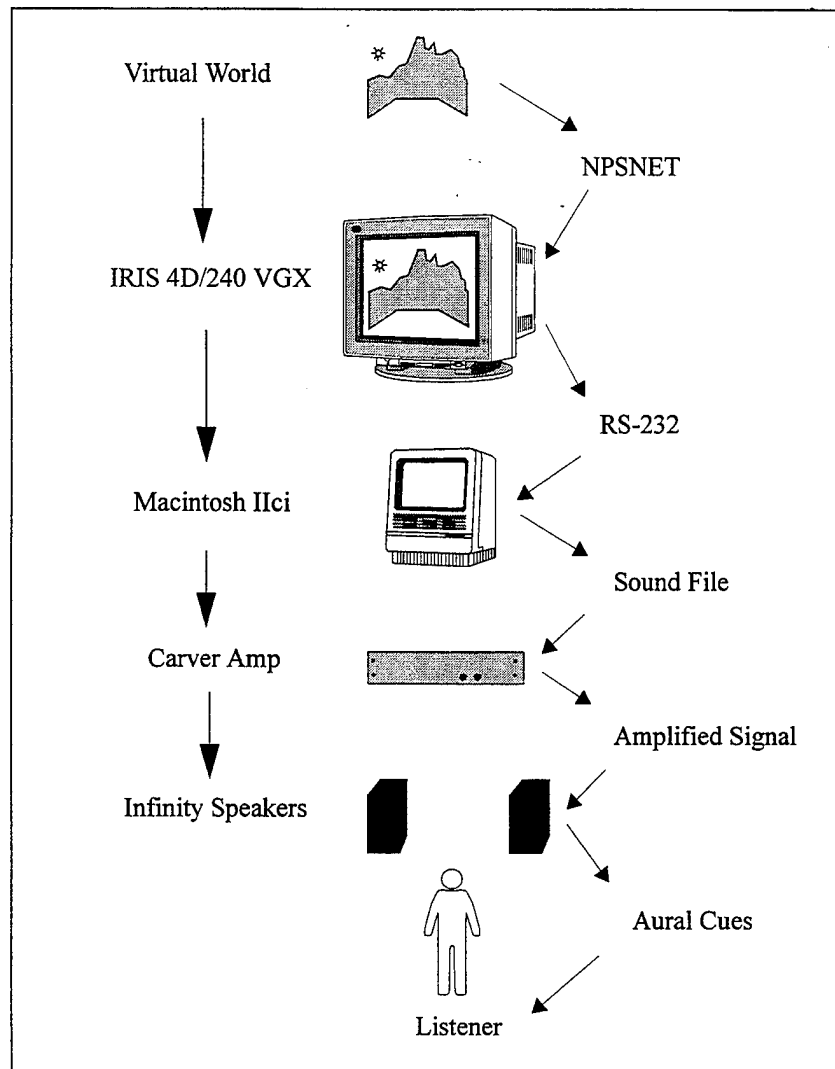
Figure 1: Overview of NPS-Sound.

4. Problems

In order to play an audio file in real-time, the file had to be stored as a resource file in the system folder on the MAC. As a result, only small audio files could be played because of the size limitation of the system folder. Too much time was also wasted by the *FontesTalk II* program in searching the system folder in order to decipher which audio file to play. Only discrete/static sounds (such as explosions) were generated for there were

problems generating continuous sounds (such as a helicopter flying overhead) as a result of the "open serial Port" XPrim in *Prograph.*

5. Conclusions

This sound system, although fairly capable, was merely a trial run in testing whether or not it was actually feasible to present aural cues in real-time to users of NPSNET. The result was that the aural cues did in fact increase the level of immersion of NPSNET users. The trials and tribulations of this research effort validated the use of aural cues for use in NPSNET and forged the permanent foundation for future NPSNET sound servers.

C.   NPSNET SOUND SERVER

From September 1991 to September 1992, the second attempt to add aural cues to NPSNET was conducted by Leif Dahl. As a Master of Computer Science student under the direction of his thesis advisors, Michael Zyda and David Pratt, Leif Dahls' efforts in adding sound to NPSNET culminated in his Master's Thesis: *NPSNET: Aural Cues for Virtual World Immersion* [DAHL92]. Also working with Leif Dahl during this time period was Susannah Bloch, a temporary summer hire working in the Graphics and Video Laboratory. Bloch's assistance in this research proved instrumental in achieving a successful sound system for NPSNET. Since the results of this research are documented in Dahl's Thesis, there is no need to restate the hardware and software specifics. However, a general overview follows.

1. General Overview

Many changes were made from the original sound system. The MAC was taken out of the real-time sound generating loop and was replaced by the EMAX II 16 Bit Digital Sound System [EMU89]. The MAC was then used off-line to control the functions of sound creation, modification, sampling, and storage. A *Sound Accelerator* digital audio card was added to the MAC and used in conjunction with the Analog-to-Digital (AD) converter of *Sound Designer II* [DIGI90]. The interface between NPSNET and the sound system was

now accomplished through an IRIS Indigo Elan and the EMAX II. The interface was established via an Apple MIDI Interface from the RS-422 serial port on the Indigo Elan to the *MIDI IN* port on the EMAX II. This is perhaps the greatest contribution of Dahl and Bloch for now all generated sounds were controlled via the MIDI protocol [INTE83]. A *C* program on the Indigo Elan analyzes NPSNET user actions via message packets over the Local Area Network (LAN). If a certain user action has a sound associated with it, a series of MIDI commands are sent to the EMAX II. The EMAX II deciphers the MIDI commands and generates the appropriate sound. This sound signal is then routed to the Carver power amplifier for output to the two Infinity speakers which generate the appropriate aural cues. See Figure 2 for an overview of the NPSNET Sound Server.

2. Conclusions

Establishing the MIDI interface between the Indigo Elan and the EMAX II increased the range of audio possibilities for use in NPSNET due to the immense amount of flexibility associated with the MIDI protocol. However, no dynamic/moving sounds were presented, for the emphasis was on creating the MIDI interface and generating static sounds such as rifle fire and explosions. But most important, as in the first sound system, the addition of aural cues still continued to increase the level of immersion of the NPSNET player, and as a result warranted further research and development.

D.   NPSNET-PAS

From September 1992 to September 1994, another Master of Computer Science student from NPS, John Roesli, under the direction of his thesis advisors Michael Zyda and John Falby, studied ways to enhance the current MIDI-based sound server for NPSNET. John Roesli's research efforts culminated in his Master's Thesis: *Free-field Spatialized Aural Cues for Synthetic Environments* [ROES94], in which a new MIDI-based sound system was developed for integrating aural cues into NPSNET. This new sound system was called NPSNET-Polyphonic Audio Spatializer (NPSNET-PAS). Again, since the results of
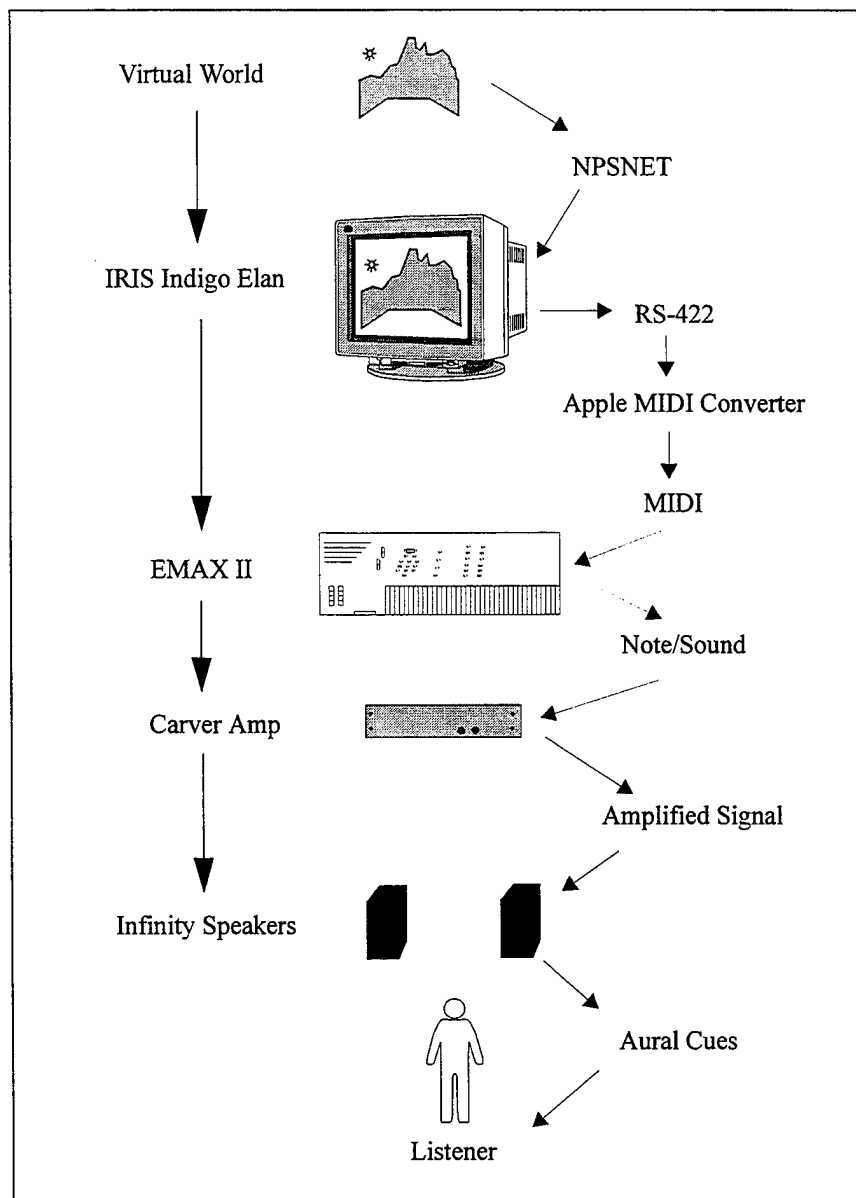
Figure 2: Overview of NPSNET Sound Server.

this research are documented in Roesli's thesis, there is no need to restate the hardware and software specifics. However, a general overview is again provided.

1. General Overview

The primary goal of Roesli's thesis was to enhance the effectiveness of the aural cues by spatializing these cues into two dimensions. The same MIDI interface between

NPSNET and the sound system was utilized. The functionality of the sound server software was enhanced and additional sound equipment was procured. Specifically, two additional speakers were added to the existing sound system so that the listener could be surrounded by a quad configuration of speakers. A subwoofer processor and a pair of subwoofers were added to generate very low frequencies around the listener. A mixing board was also added to control the levels of all audio signals. See Figure 3 for an overview of NPSNET-PAS.

2. Conclusions

The goal of Roesli's thesis was realized, for NPSNET-PAS did in fact produce spatialized aural cues in two dimensions for use in NPSNET. Furthermore, the addition of the subwoofers dramatically added to the realism of the aural cues. During NPSNET demonstrations, numerous participants commented that the low frequencies generated by the subwoofers dramatically increased their immersion into the virtual environment of NPSNET. Again, as in the previous sound systems, no dynamic/moving sounds were presented. However, the MIDI pitch bend command was implemented to coincide with the host machine's vehicle speed in an effort to increase the overall realism of the vehicle's sound. As a result, when the vehicle's speed increased or decreased, the vehicle's pitch correspondingly increased or decreased. NPSNET-PAS, the third generation of NPSNET sound systems, has provided the greatest level of immersion for players in NPSNET thus far, and set the foundation for spatializing aural cues in three dimensions.
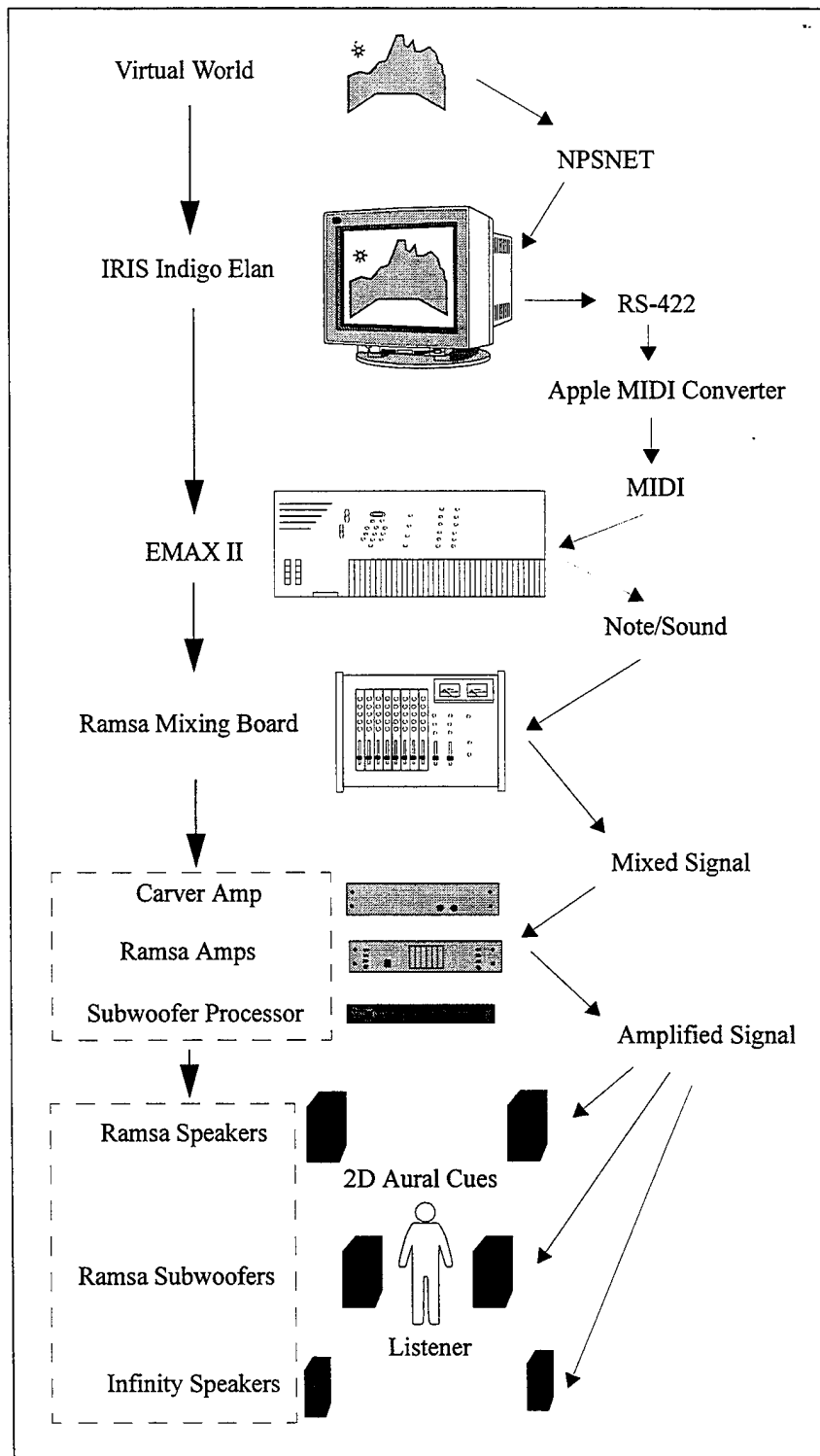
Figure 3: Overview of NPSNET-PAS.

III. BACKGROUND

In order to better understand the concept of 3D sound and how it can be used in a virtual environment application, a brief background is presented in the following areas: wave properties of sound, 3D sound perception, Inverse-Square Law, and MIDI.

A.   WAVE PROPERTIES OF SOUND

Sound, like light, has properties of waves. These wave properties are summarized as follows [WILL76]:

- Propagation: continuous waves traveling in a uniform medium propagate in straight lines perpendicular to the advancing wavefronts.
- Reflection: occurs when a wave is turned back (reflected) upon encountering a barrier that is the boundary of the medium in which the wave is traveling.
- Refraction: is the bending of the path of a wave disturbance as it passes obliquely from one medium into another of different propagation speed.
- Interference: can be constructive (see Figure 4) or destructive and is based on the principle of superposition which in terms of sound is as follows:
  -- ...the same portion of a medium can simultaneously transmit any number of different sound waves with no adverse mutual effects. If several sound waves travel simultaneously through a given region of the air medium, air particles in that region will respond to the vectorial sum of the required displacements of each wave system. [EVER91a]
- Diffraction: the spreading of a wave disturbance beyond the edge of a barrier.

In working with sound, one must have a good understanding of these wave properties. It is through these properties that we describe the occurrence of most common types of sound phenomena. For example, tap a tuning fork and listen to the generated tone. Then, slowly turn the tuning fork in your hand. You will hear louder and softer tones as you turn the tuning fork. Why are there louder and softer tones? The reason is based on the property of interference. The soft tones are from the original tapping of the tuning fork. The loud tones are caused by the constructive interference of the original two sound waves

which only became apparent when moving the tuning fork. Figure 4 depicts this example of the property of interference.
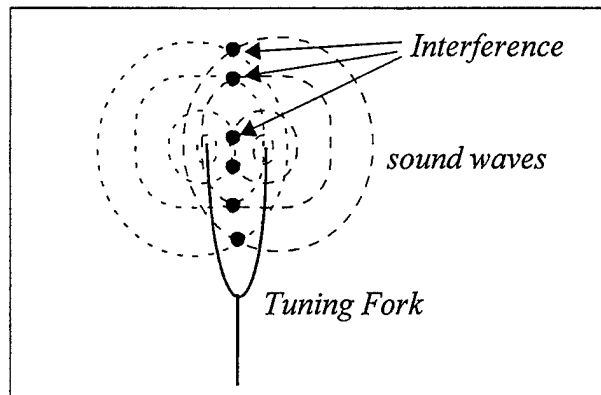
Figure 4: Interference of Sound Waves. After [GILL95b].

Another example can be found with loudspeakers. Why does sound propagate spherically from a loudspeaker? One reason is based on the property of diffraction. Exactly how a sound wave is diffracted is dependent upon the wavelength of the sound source and the size of the aperture. See Figure 5 for a depiction of how the property of diffraction works.
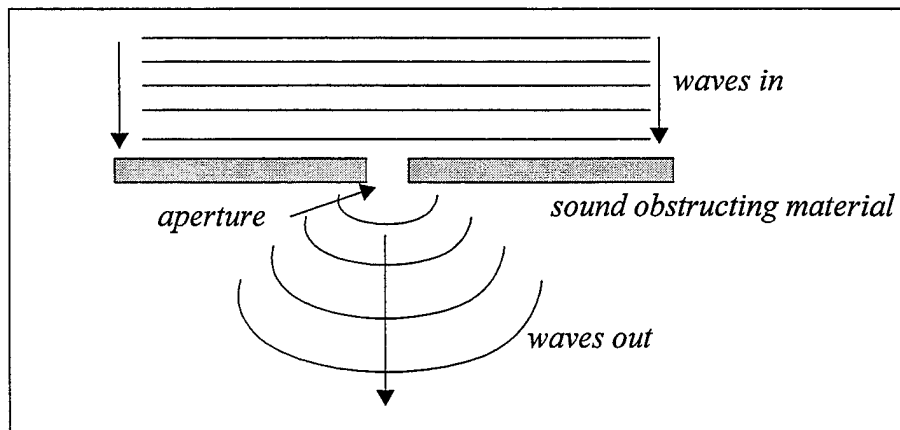
Figure 5: Diffraction of Sound Waves. After [EVER91a].

B.   3D SOUND PERCEPTION

To understand the concept of 3D sound perception, a discussion of psychoacoustics, sound localization, the Duplex Theory, the head-centered coordinate system, and the precedence effect is presented.

1. Psychoacoustics

Recording sound is fairly simple, but evaluating sound is not. The difficulty is that sound cannot be measured solely as a physical quantity, for attached to the physical nature of sound are psychophysical qualities. "Measuring these psychophysical qualities includes mental processing, and can only indicate probabilities of human response to a stimulus" [BEGA94]. Thus, to measure sound we must keep in mind how the sound is perceived. The psychophysics of sound is termed psychoacoustics and plays a crucial role in determining how we humans spatialize sound. As a result, the effectiveness of any type of sound delivery system stems primarily from the psychoacoustic nature of sound. In other words, no matter how good a sound system might be in terms of its accuracy to physical laws, the bottom line in evaluating a sound delivery system comes from how good it is perceived to be. (A great source which illustrates much of the way we humans perceive sound is a book titled *Auditory Scene Analysis* by A. Bregman [BREG90].)

2. Sound Localization

How we humans localize sound is still a very active area of research. Even after years of research, we still do not know exactly how we localize sound. What we do know is that we humans use certain localization cues to help us distinguish sounds. These localization cues include: interaural time difference, interaural intensity difference, pinna response, shoulder echo, head motion, early echo response, reverberation, and vision [TONN94]. Still, there are other cues such as atmospheric absorption, bone conduction, and a listener's prior knowledge of the sound source [ERIC93]. As research in this field continues, the list of localization cues, and the theories behind these cues, will no doubt

continue to grow. See APPENDIX H: SOUND PERCEPTION EXPERIMENTS on page 167 for some experiments involving sound localization. To help explain why there exists so many theories, one needs to look at the multiple acoustic paths (see Figure 6) that a sound source travels before it reaches our eardrum. Some of these various paths include:
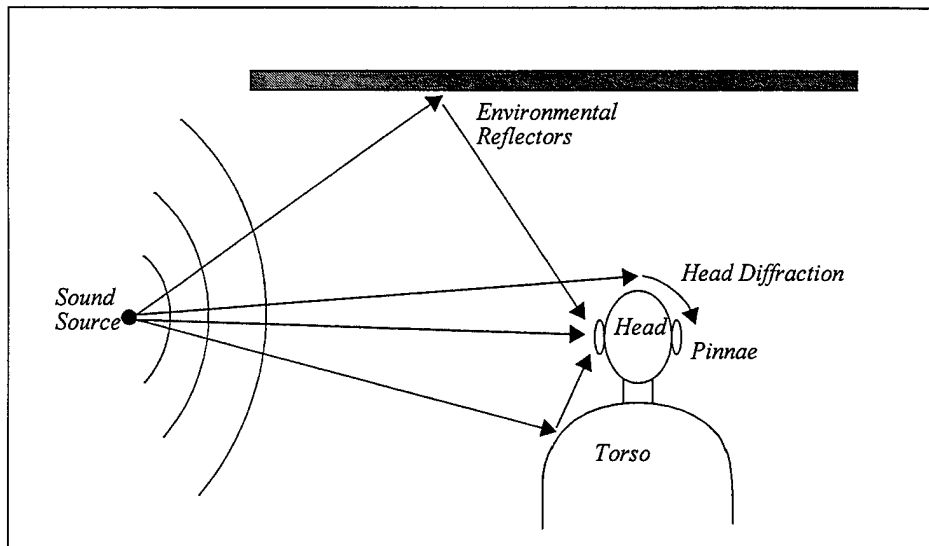
Figure 6: Acoustic Paths. From [DUDA95].

environmental reflectors, head diffraction, the head itself, pinnae, and torso.

a. The Pinnae

New studies are revealing that the outer ears (the pinnae) play a much larger role in sound localization [WENZ92] [BEGA94]. Numerous experiments have shown that the shape of the pinnae (pinnae is plural and pinna is singular) provides for a spectral shaping of sound which is highly directional dependent [SHAW74]. Consequently, the absence of such spectral shaping severely degrades localization correctness [GARD73]. These highly directional audio cues provided by the pinnae's spectral shaping are chiefly responsible for producing the perception known as externalization -- the outside-the-head sensation [PLEN74].

b. The Duplex Theory

The Duplex Theory, formalized by Lord Rayleigh in 1907, suggests that the head itself provides the listener with two localization cues [LORD07]. One cue is the Interaural Time Difference (ITD), which is the time delay experienced when a sound reaches one ear before the other. The other cue is the Interaural Intensity Difference (IID), which is the intensity difference between the two ears as a result of head diffraction. These two cues are depicted in Figure 7.
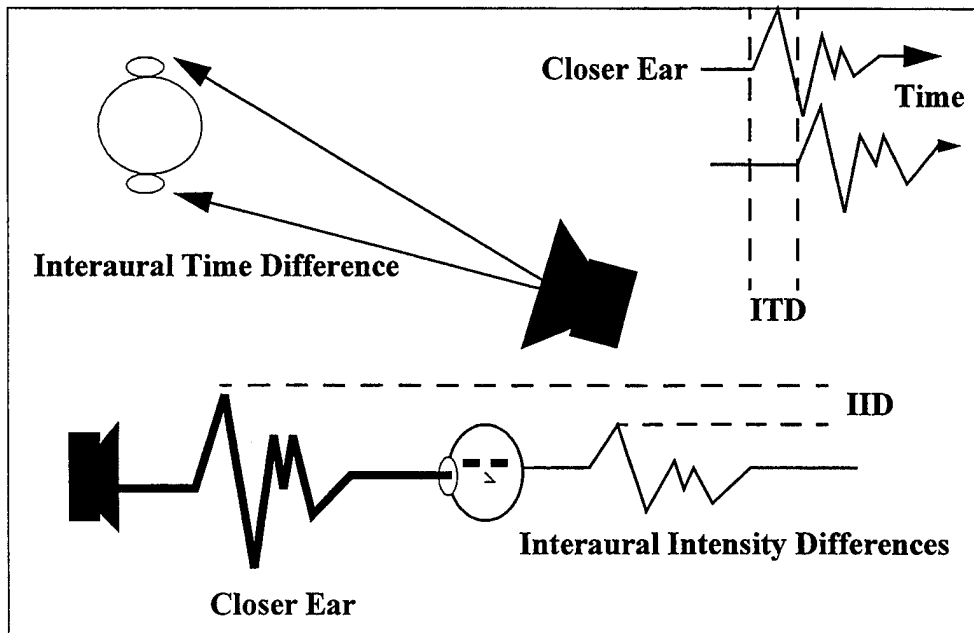
Figure 7: Two primary cues of sound localization [WENZ90].

3. Head-Centered Coordinate System.

Because the head gives us the ITD and IID cues as described in the Duplex Theory, any coordinate system used to model how a listener localizes a sound should place the middle of the head at the center of the coordinate system. Figure 8 represents this head-centered coordinate system. The elevation is represented by ϕ and is determined by such cues as pinnae reflections and torso diffraction. The azimuth is represented by θ and is determined by the ITD and IID cues where θ is estimated by the ITD at low frequencies
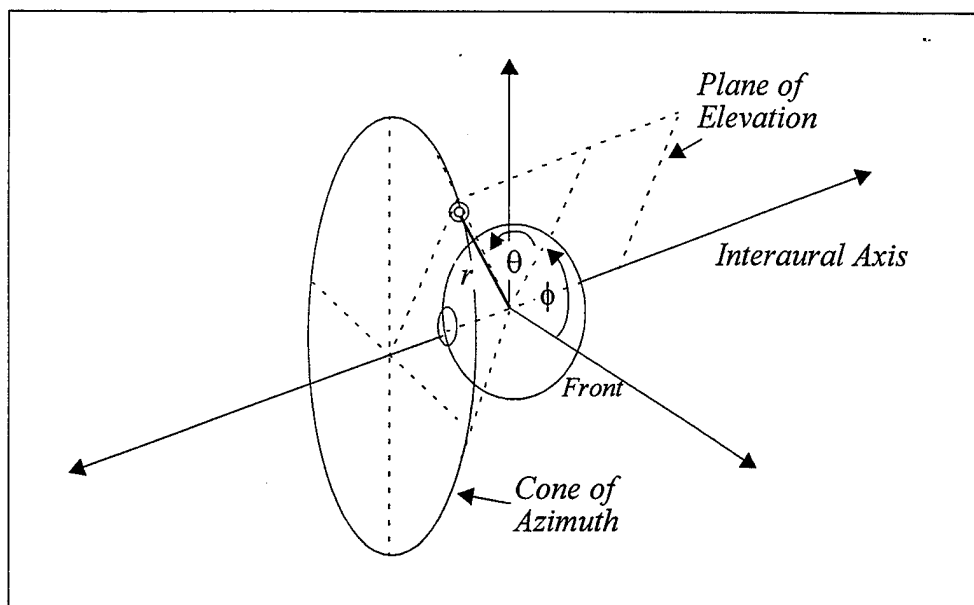
Figure 8: Head-Centered Coordinate System. From [DUDA95].

(below 1500 Hz) and θ is estimated by the IID at high frequencies (above 1500 Hz). The range (distance to the sound source) is represented by *r*, and is determined by such cues as intensity, direct/reverberant ratio, and head motion. [DUDA95]

By establishing this head-centered coordinate system, we now have a basis for which mathematics can be used to derive the ITD and IID cues as described in the Duplex Theory. For example, given the following equation:

$$\lambda = \frac{c}{f} \qquad\qquad \text{Eq 1}$$

where,

λ is the wavelength,

f is frequency,

c is the speed of light.

We can now derive both the ITD and the IID based on the azimuthal angle θ as shown in Figure 9.
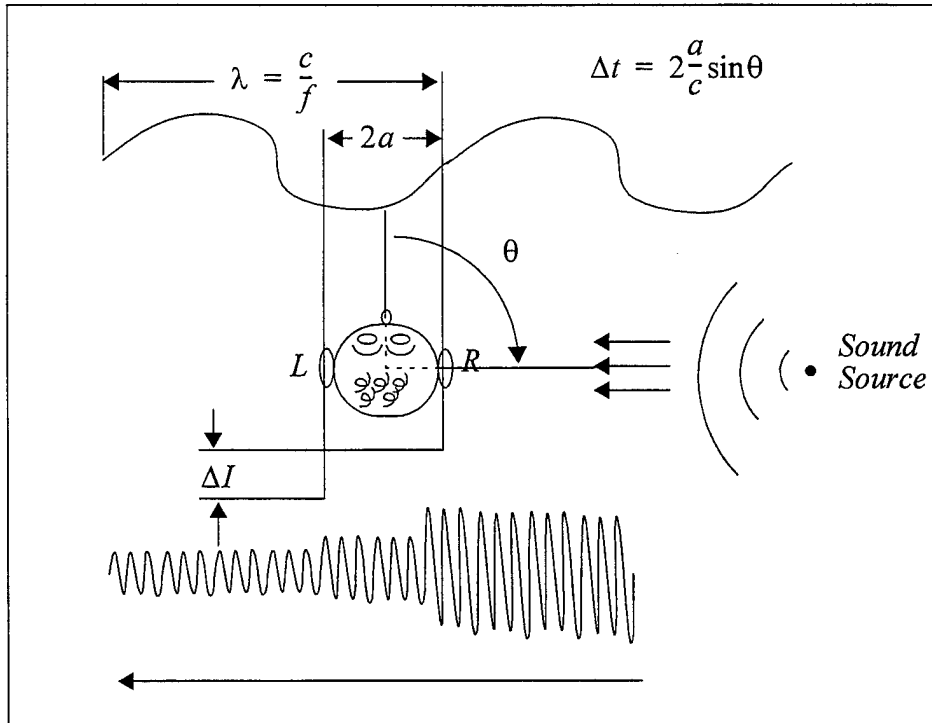
Figure 9: Mathematics of the Duplex Theory. From [DUDA95].

A good rule of thumb is that, on average, there is a millisecond delay (the ITD) between the hearing of our both our outer ears as shown in see Figure 10 [GILL95a].
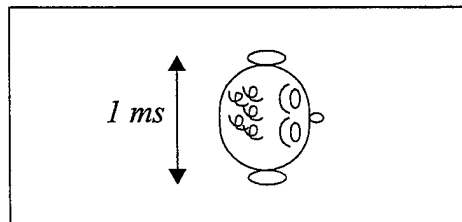
Figure 10: Approximate ITD.

This is the foundation of using the Head-Related Transfer Function (HRTF) to reproduce the delay between our ears using a headphone sound delivery system. A more in depth discussion of the HRTF is presented in Chapter V. HEADPHONES VS. FREE-FIELD DELIVERY SYSTEMS.

4. The Precedence Effect

Another cue which can both aid and hinder our ability to localize sounds is based upon the Precedence Effect (PE). The PE means that when and where we perceive the sound first will influence the direction from which we think the sound source is emanating (see Figure 11). This helps us to distinguish an original sound source from that of its echoes.
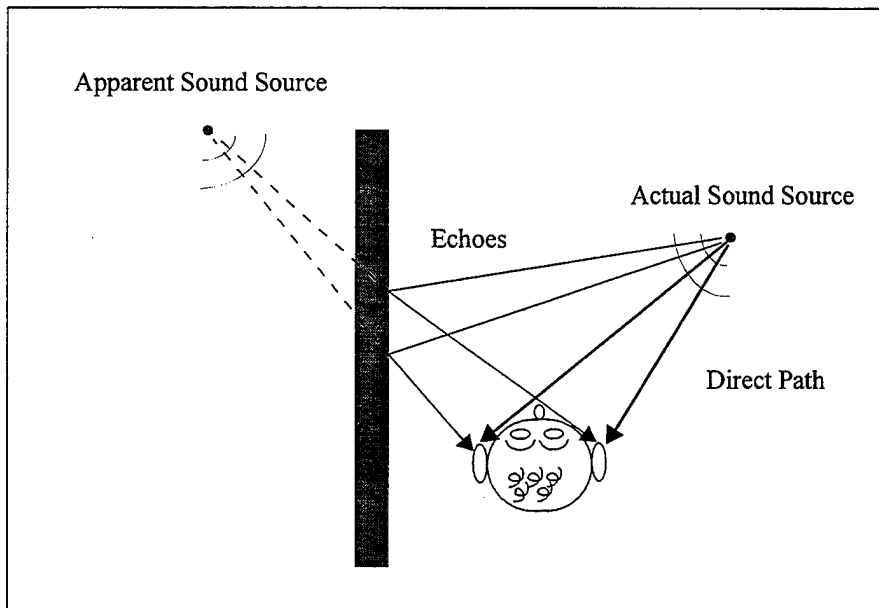
Figure 11: The Precedence Effect. From [DUDA95].

In looking at Figure 11, since the direct path of the actual sound source arrives at our ears first, we believe the sound is coming from the actual sound source. Thus, based on the PE, we have correctly localized the sound source. However, if instead we first had heard the sound coming from the path of the echoes, we would think that the sound was coming from the apparent sound source as opposed to the actual sound source. So now, based on the PE, we have incorrectly localized the actual sound source. As can be seen, the PE gives us another cue with which to localize sound. The PE is also called The Law of the First Wavefront.

C.   THE DECIBEL

The *bel* (named after Alexander Graham Bell) is defined as the logarithm (to the base 10) of the ratio of two powers as shown in Eq 2 [EVER91a].

$$L \, (bels) \;=\; \log \frac{W_1}{W_2} \qquad\qquad \text{Eq 2}$$

where,

L is the level measured in *bels*,

W_1 and W_2 are measurements in Power.

The *bel*, however, is too large for working with sounds, so the *decibel* (1/10th of a *bel*) was adopted as shown in Eq 3 [EVER91a].

$$L \, (decibels) \;=\; 10 \log \frac{W_1}{W_2} \qquad\qquad \text{Eq 3}$$

In looking at Eq 3, we see that the decibel (dB) is a ratio and must be used in reference to something. The standard value used as this reference is derived from the lowest threshold of hearing which is equal to 10^{-12} W/m^2 [SAPP95]. This value is known as the reference energy and is sometimes referred to as 0 dB. This is the lowest sound pressure level that we humans can hear. If a sound source had an energy of 10^{-9}W/m^2, then we would do the following to calculate it's decibel level:

$$10 \log \frac{10^{-9}}{10^{-12}} \;=\; 10 \log 1000 \;=\; 30 dB \qquad\qquad \text{Eq 4}$$

Another common use of dB is to establish a reference point in order to adjust the gain on numerous types of sound systems. In this case, a dB_v is equal to 1 volt. This scale is used to determine the positive or negative gain relative to the optimal signal level for a particular sound system. As a result, a level of 0dB is equal to the sound system's optimal signal level. Thus, a positive or negative gain relates to positive or negative levels from the particular sound system's optimal signal level. [SAPP95]

25

D. INVERSE-SQUARE LAW

The following is a summary of the Inverse-Square Law and it's derivation taken from the *Handbook for Sound Engineers* [EVER91a].

The Inverse-Square Law can only be applied to sound in a free field. The Inverse-Square Law states that the intensity of sound is inversely proportional to the square of the distance from the source. But what is sound intensity? Sound Intensity is defined as the sound power per square centimeter (W/cm2). Thus we have the following:

$$I = \frac{W}{4\pi r^2}$$

Eq 5

where I is the sound intensity in W/cm^2,

W is the sound power of the source in watts,

and r is the distance from the source in cm.
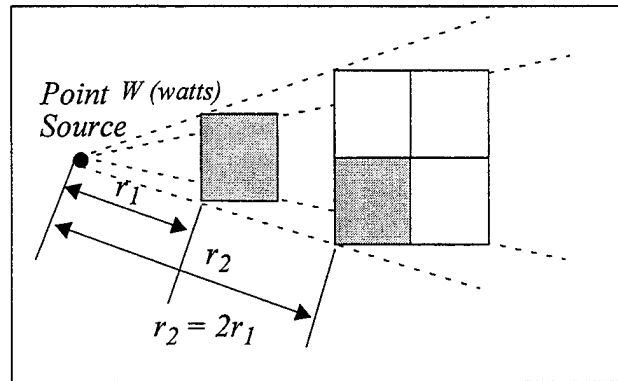
Figure 12: Inverse-Square Law. After [EVER91a].

In Figure 12, a sound source is emanating in free space flowing outward. At a distance r_1 from the source we have the following:

$$W = I_1 \times 4\pi r_1^2$$

Eq 6

And, at a distance r_2 from the source we get:

$$W = I_2 \times 4\pi r_2^2$$

Eq 7

Since the watts, W, at either distance is the same, we can set Eq 6 and Eq 7 together and get the following:

$$I_1 \times 4\pi r_1^2 = I_2 \times 4\pi r_2^2 \qquad\qquad \text{Eq 8}$$

Eq 8 can be then rewritten as:

$$\frac{I_1}{I_2} = \frac{4\pi r_2^2}{4\pi r_1^2} = \frac{r_2^2}{r_1^2} \qquad\qquad \text{Eq 9}$$

Eq 9 is the Inverse-Square Law. But remember, the Inverse-Square Law is based on intensity. And, intensity is a difficult parameter to measure requiring special techniques. Sound pressure, on the other hand, is an easily measured parameter based on the decibel as described above. The question now is how to express the Inverse-Square Law in terms of sound pressure? The intensity at r_2 is one-forth that at r_1. Since sound pressure is proportional to the square root of the intensity, the sound pressure at r_2 is one-half that at r_1 (i.e $\sqrt{1/4} = 1/2$). Thus, remembering that a decibel is always a ratio, a drop of 1/2 corresponds to a drop of 6 dB. Therefore, in the free field, sound pressure drops off at the rate of 6dB for distance doubled.

A very important point to keep in mind is that the decibel applies only to power-like quantities. Thus, acoustic intensity, which is power per unit area in a specific direction, can be expressed (and is expressed) in decibels. However, when sound is measured, it is normally measured as a sound pressure, not as an acoustic power. But the square of this typically measured sound pressure remains proportional to acoustic power. So, the important thing to remember is that when acoustic power is being compared the following formula must be used:

$$L\,(decibels) = 10\log\frac{Pressure_1^2}{Pressure_2^2} \qquad\qquad \text{Eq 10}$$

However, when sound pressure is being compared, the following formula must be used:

27

$$L\,(decibels)\ =\ 20\log\frac{Pressure_1}{Pressure_2}$$

<div align="right">Eq 11</div>

Therefore, the 10 log is used for power ratios, and the 20 log is used for sound pressures.

This concludes the summary taken from the *Handbook for Sound Engineers* [EVER91a].

E.   MIDI

The Musical Instrument Digital Interface (MIDI) is a standardized communication protocol. It was developed by researchers in Japan and was first released as MIDI Specification 1.0 in 1983 [INTE83]. Its purpose was to establish a communication standard for which electronic musical instruments could effectively communicate in both real-time and nonreal-time. It is important to note that MIDI does not transmit any sound/audio data. It just facilitates communication among the attached MIDI capable devices.

1. Hardware Structure

MIDI communication is made possible through a MIDI cable and the MIDI In, MIDI Out, and MIDI Thru ports on the MIDI devices. The MIDI cable consists of a shielded, twisted pair of conductor wires having a male 5-pin Deutsche Industri Norm (DIN) on either end of the cable. This cable allows for asynchronous serial communication at the rate of 31.25 Kbaud (+/- 1%). However, the MIDI ports are unidirectional and only allow communication to one direction. The reason for this one way communication is that the MIDI In port only allows incoming information, and the MIDI Out port only allows outgoing information. The MIDI Thru port duplicates the information received by the MIDI In port and sends this information out the MIDI Thru port. The MIDI Thru port is typically used for daisy chaining multiple MIDI devices.

2. Communication Format

Communication in MIDI is accomplished through the following five types of MIDI messages along with their associated data:

- Channel Voice
- Channel Mode
- System Common
- System Real-Time
- System Exclusive

These five messages are described in Figure 13. Furthermore, these messages can be sent

		Voice	used to control the voices of a device
M E S S A G E S	**C H A N N E L**	Voice	used to control the voices of a device
		Mode	used to determine how a device responds to voice messages
	S Y S T E M	Common	messages intended for all devices
		Real-Time	used to synchronize the devices
		Exclusive	messages that allow manufacturers to implement functions unique to their devices
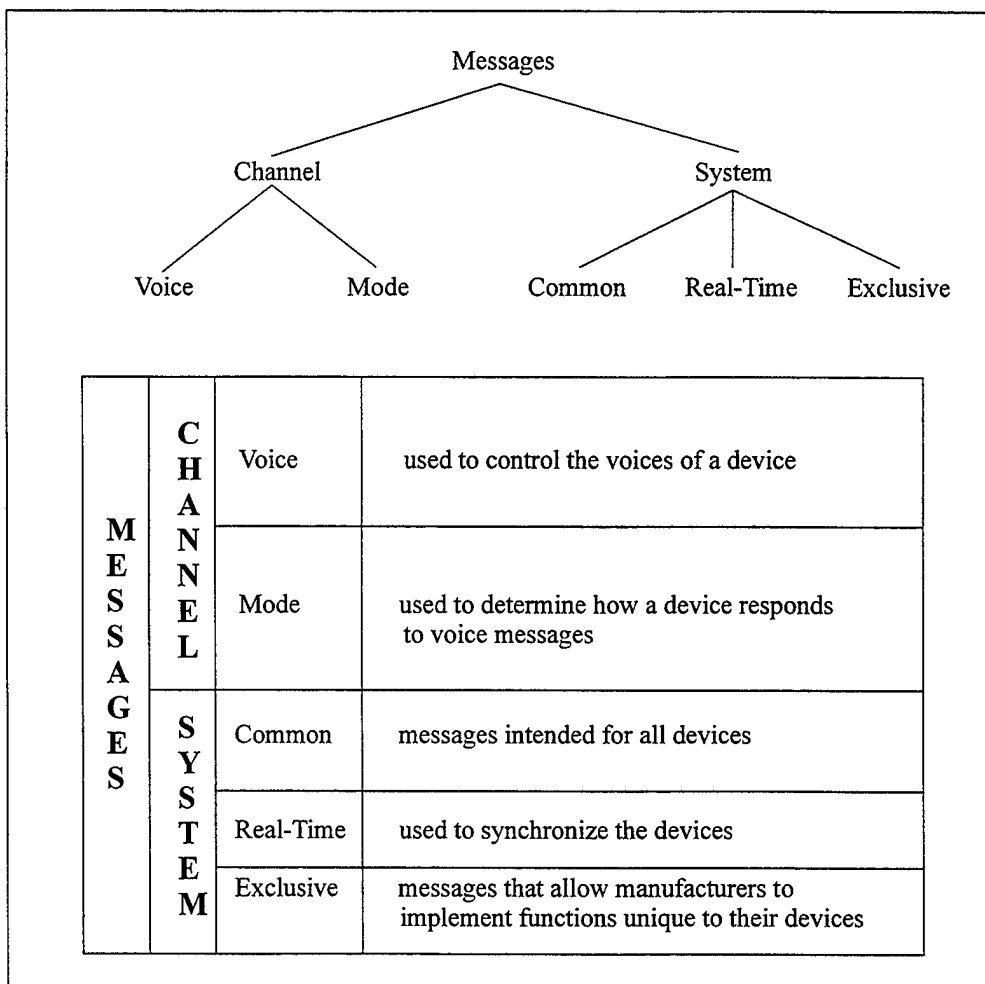
Figure 13: Structure of MIDI messages. From [DOAN94].

on any one or all of sixteen possible independent channels. In turn, a MIDI device can be assigned any one channel or any combination of up to sixteen channels to receive these messages.

Although behind the technical power curve, MIDI is still in use with today's sophisticated computers and electronic musical equipment. However, improvements are warranted, such as the ZIPI Music Parameter Description Language [MCMI94]. But for now, MIDI continues to be used world wide and in numerous applications.

IV. THE AUDITORY CHANNEL

The spatialization of sound through applications of 3D sound perception improves the level of immersion for the listener within a virtual environment (VE) and is known as virtual audio. This spatialized sound application has come to fruition because, "the fact that audio in the real world is heard spatially is the initial impetus for including this aspect within a simulation scenario" [BEGA94]. As a result, "Virtual audio is the perception of being immersed in a listening environment different from the actual one in which a listener is physically located" [ERIC93]. Thus, "the goal of virtual audio technology is to create the illusion that a listener is in a particular acoustic environment" [ERIC93]. The National Academy of Science's Committee on Virtual Reality Research and Development, however, refers to virtual audio as the *Auditory Channel* in a Synthetic Environment (SE). (Synthetic Environment is the term chosen by the Committee on Virtual Reality Research and Development to represent all of the following types of systems: virtual reality, cyberspace, virtual environments, teleoperation, telerobotics, and augmented reality [DURL95].) The term auditory channel is noteworthy for it complements the Committee's term for the visual interface into a SE, the *Visual Channel*. Thus, the auditory channel is no longer an afterthought, but rather an integral part of a SE.

A. 3D AUDITORY DISPLAYS

An auditory display is the vehicle by which audio cues are presented to the listener through the auditory channel in a SE. These displays include:

- Audification, in which the acoustic stimulus involves direct playback of data samples, using frequency shifting, if necessary, to bring the signals into auditory frequency range. [DURL95]
- Sonification, in which the data are used to control various parameters of a sound generator in a manner designed to provide the listener with information about the controlling data. [DURL95]

31

If this sounds a bit confusing, it might be helpful to compare an auditory display with a visual display. For example, when one looks at a visual display on a monitor, one sees a visual image comprised of various colored pixels. Conversely, when one hears an auditory display, one hears an auditory image comprised of various generated sounds. In summary, "the combination of 3-D sound within a human interface along with a system for managing acoustic input is termed a 3-D auditory display" [BEGA94].

B. EXTERNALIZATION

Externalization occurs when a listener perceives an auditory image outside the listener's head. Conversely, when someone is listening to a conventional stereo recording through headphones, the auditory image is located inside the listener's head. This is called internalization. However, it is externalization that plays a critical role in the auditory channel. It should be noted that an auditory image is not the same as an acoustical image. "Auditory events have apparent locations in auditory space. Acoustical events have actual locations in the physical space surrounding the listener" [MART92]. Thus, psychoacoustics plays a much greater role in determining and evaluating auditory images as opposed to acoustical images.

C. SPATIALIZATION

When an externalized auditory image, along with various localization cues, is combined with a certain azimuth and elevation, a spatialized auditory image is formed. Again, psychoacoustics plays a critical role, "because the perception of the spatial properties of a sound field is an important component of the overall perception of real sound fields" [DURL95]. Thus, the level of one's immersion in a VE is directly proportional to how well the spatialized auditory image conforms to the listener's perception of its real-world counterpart.

D.   SEMANTICS

Because the use of audio in VEs is a relatively new area of research, some of the terminology used so far may seem a bit confusing. Nevertheless, various researchers use slightly different names to say pretty much the same thing. For example, the concept of 3D sound has been described by various researchers as spatialized audio, spatial audio, virtual acoustics, virtual audio, 3-D auditory display, 3D spatial audio, auditory images, virtual auditory images, binaural audio, binaural acoustics, auditory localization, spatialized sound, spatial sound, spatial image, auditory channel, and some others. Some of these terms are indeed identical concepts, but others are not; hence the confusion. Furthermore, the semantics of these terms varies with different applications. Hopefully, in the near future, some form of standardization will be placed on the terminology of 3D sound. Perhaps the National Academy of Science's Committee on Virtual Reality Research and Development could help to implement some standardization on the terminology of 3D sound as it pertains to VEs. If so, the inherent complexity 3D of sound would at least be a little less confusing.

E.   INTERFACE DEVICES

There are two primary interface devices for generating 3D sound within a VE: headphones and loudspeakers. Each device has its advantages and disadvantages, and each device is actively being researched within the virtual reality community. It should be noted that it is not the actual devices themselves that are being researched, but rather how the devices should be utilized.

In other words, from the viewpoint of synthetic environment (SE) systems, there is no need for research and development on these devices and no need to consider the characteristics of the peripheral auditory system to which such devices must be matched. What is needed, however, is better understanding of what sounds should be presented using these devices and how these sounds should be generated. [DURL95]

The next chapter discusses the advantages and disadvantages of using headphones and loudspeakers for generating 3D sound within a VE.

V. HEADPHONES VS. FREE-FIELD DELIVERY SYSTEMS

There are numerous applications in the real world which include 3D sound. Some of these applications include [BEGA94]:

- Improving the quality and ease of interaction within a human interface.
- Improving situational awareness by providing an extra channel of feedback for actions and situations both in and out of view of the listener.
- Reducing stress caused by communication overload in the modern airline cockpit.
- Improving sound quality in movie theaters (not the same as surround sound).
- Improving the level of immersion in virtual environments.

In evaluating the two types of sound delivery systems (headphones or free-field), it is important to consider its associated application. For the evaluation to be consistent, it is not appropriate to mix applications between the two types of delivery systems. For example, it is not valid to compare a headphone sound system for reducing stress caused by communication overload in the modern airline cockpit with a free-field sound system for improving the sound quality in movie theaters. Thus, the merits of each delivery system are directly related to the specific type of application utilized. Accordingly, the focus of this research is to evaluate the advantages and disadvantages of headphones and free-field systems in the application of improving the level of immersion in VEs.

A.   HEADPHONE DELIVERY SYSTEMS

The type of headphones used in virtual environments (VEs) is essentially the same type used for listening to one's stereo system. These headphones come in all types of shapes and sizes. However, "most users of 3-D sound systems will use either supraaural (on the ear) or circumaural (around the ear) headsets" [BEGA94]. There are advantages and disadvantages to both types of systems. Supraaural headsets are nice because it is easy to communicate with whomever is wearing the headsets, for the listener's ears are not completely covered. Conversely, to effectively communicate with someone wearing circumaural headsets, one would have to talk into a microphone which was integrated into

the listener's sound system. On the other hand, because circumaural headsets cover the entire ear:

> speaker diaphragms with better frequency responses can be used, greater isolation from extraneous noise can be achieved, and better, more consistent coupling between the ear and the headset is insured. [BEGA94]

Regardless of which type of headphone is used, a binaural reproduction of sound must be reproduced and is based on the Head-Related Transfer Function (HRTF).

1. Head-Related Transfer Function

A method of recreating the perception known as externalization, provided by the spectral shaping of the pinnae, is to capture the sum of all aspects affecting localization by the pinnae into a filter that can be applied to a sound. The aspects affecting localization can be captured by placing tiny microphones in a listener's ears, referred to as biaural recording, and producing a short sound pulse (see APPENDIX G: BINAURAL RECORDINGS). The output of the microphones can be measured and used to create such a filter. The advantage to this method is that it captures the aggregate spatial cues for a particular source location, listener, and environment. These filters are called finite impulse responses (FIR) and are referred to as a the HRTF. In other words, "The spectral filtering of a sound source before it reaches the ear drum that is caused primarily by the outer ear is termed the head-related transfer function (HRTF)" [BEGA94]. By applying this filter to a given sound source, the spatial location of the original filter can be recreated [WENZ90]. In summary, "The HRTF is a linear function that is based on the sound source's position and takes into account many of the [localization] cues humans use to localize sounds..." [TONN94].

2. Advantages

Perhaps the greatest advantage of using headphones over loudspeakers is that "they fix the geometric relationship between the physical sound sources (the headphone drivers) and the ears" [BURG92]. Thus, when used in conjunction with a head tracker such as a

Polhemus Fastrack, the listener's head position can be continually monitored. As a result, when the listener turns his head, the directionality of the listener's perceived sound, which is generated through the headphones, correspondingly changes in relation to the listener's head movement. This head movement correlation is extremely important for it "can allow a listener to improve localization ability on the basis of the comparison of interaural cues over time" [BEGA94]. Furthermore, when used in conjunction with a visual cue, the listener can better approximate its spatial location. This audio and visual association is known as the ventriloquism effect or the visual capture effect.

Another advantage of using headphones is that they are individualistic devices. A listener can be immersed in his own VE without being distracted by sounds from another listener's perspective in the same or entirely different VE. Conversely, a listener using headphones will not disturb the privacy of anyone in close proximity.

Cost is another advantage. A pair of headphones is significantly cheaper than a pair of loudspeakers. Granted, there is additional equipment needed, such as specialized digital signal processors (DSP) for generating 3D sound in real-time through a pair of headphones. But, DSP's can also be found in loudspeaker sound systems, thus headphones are relatively cheaper.

3. Disadvantages

Although HRTF filters have provided a fairly accurate model of sound localization, they are not without problems. A limited resolution of about 5 to 20 degrees, when combining both azimuth and elevation data, is about the best that has been achieved. This poor resolution is known as localization blur [BLAU83]. Furthermore, back-to-front confusion [OLDF84] and elevation confusion [WENZ92] are also present for reasons which are not yet totally understood. One explanation is the so-called cone-of-confusion [MILL72] caused by sounds emanating from certain bearings which produce the same ITDs and IIDs. In short, because of the complexities in determining how we humans perceive sound, HRTFs alone cannot provide complete spatialization of sound.

Furthermore, in order to deliver spatial audio cues via headphones, it is necessary to process enormous amounts of digital audio data. Since we only have two speakers (one for each ear), the sound must be filtered using a HRTF. Thus, the processing is extremely time consuming and cannot be performed in real-time without special hardware. One such specially designed hardware system is the Convolvotron which is a real-time sound spatializer developed by Crystal River Engineering. The Convolvotron uses a person's unique set of ear impulse responses, the Head-Related Transfer Function (HRTF), to generate the appropriate spatial sound (see Figure 14). In order to accomplish the immense
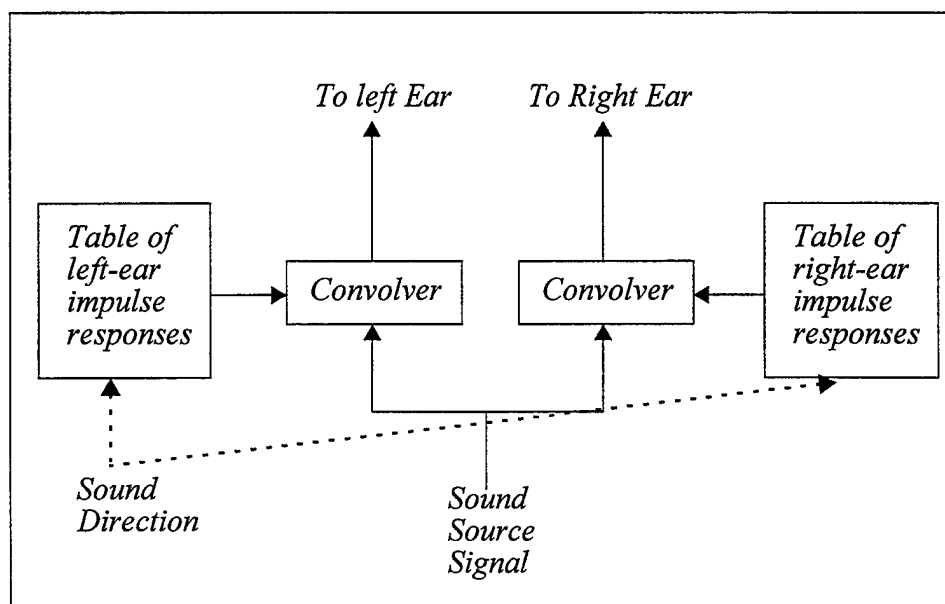
Figure 14: The Convolvotron. From [DUDA95].

amount of calculations needed in to compute spatial sound in real-time, the Convolvotron operates at an aggregate computational speed of more than 300 million multiply-accumulates per second. Figure 15 shows how the Convolvotron synthesizes spatial sound from the original input sound source. But, only four individual sound cues can be processed simultaneously. More sound cues could be added to a sound system by obtaining additional Convolvotron's, but at a price of $14,995 per Convolvotron (as of 1 January 1995), this could become prohibitively expensive.
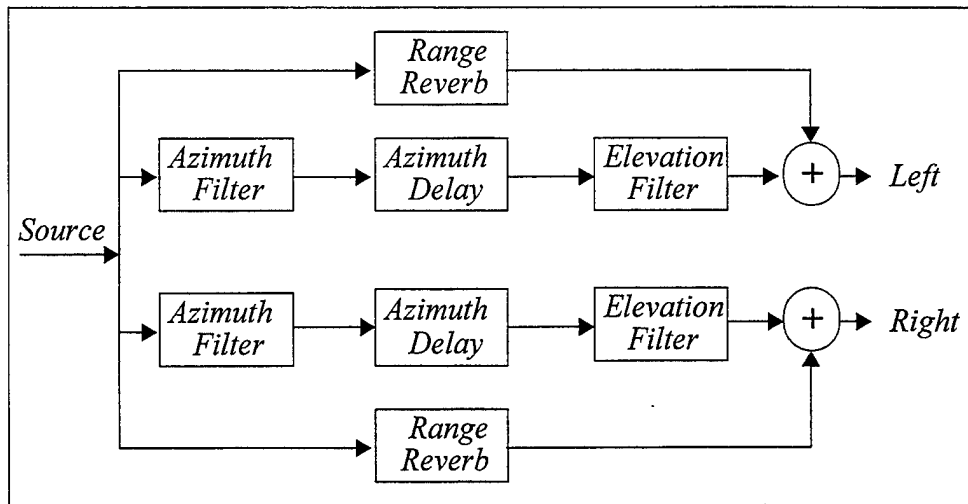
Figure 15: Synthesizing Spatial Sounds. From [DUDA95].

Other problems associated with headphones include the fact that the HRTF filters, created using the binaural recording method, are specific to the individual and as a result these filters may differ significantly from person to person. Also, the use of different types of headphones may significantly degrade effectiveness [MART92].

B.   FREE-FIELD DELIVERY SYSTEMS

A free-field delivery system gets its name from the fact that the sound is produced in the open air (i.e. free-field). Free-field systems are comprised of amplifiers and loudspeakers. The amplifier, as the name implies, simply takes an audio signal as input and amplifies it as output. The loudspeaker, in turn, receives the amplified output signal from the amplifier and generates the actual sound which is heard by the listener. As with headphone systems, there are numerous types of free-field systems which can be used for generating aural cues for use in VEs. These free-field systems are no different than one's home stereo system. In some of the more sophisticated systems, the term *studio monitor* is used instead of loudspeaker. As the name implies, studio monitors are often found in the recording studio to satisfy the most discerning ears of the record producer. Typically, a studio monitor can handle a large amount of signal power (watts) which in turn produces a

39

very *clean* sound with wide bandwidth, high dynamic range, and low distortion having a very flat response. Flat response relates to the on-axis frequency response characteristic of the monitor/loudspeaker.

There is varying opinion as to where the flat region is, but most system's aficionados will agree that a smooth, flat response from as low as possible (at least 40 Hz) to at least 5 kHz is important. Above this, opinion varies; some prefer a gradual rolloff above 5 kHz to -10 dB at 16 kHz, while other prefer a system flat to at least 10 kHz. [HENR91]

A common use of free-field systems which can enhance one's level of immersion is the use of surround sound in movie theaters. However, the term surround sound should not be confused with 3D sound. The purpose of surround sound is to surround the listener with sound -- not to spatialize the sound. For example, a typical use of surround sound in movie theaters is to have voice sounds coming from the front speakers, and to have certain sound effects played in the rear speakers. The listener is then surrounded in sound generated by the external loudspeakers. 3D sound via free-field reproduction (loudspeakers) is similar, yet very different. The goal is not to surround the listener with a somewhat arbitrary location of sound, but rather to provide the listener with the same audio cues as if the sounds were real-time actual 3D sounds and not simply sounds being generated through loudspeakers.

1. Advantages

One advantage of loudspeakers is that they do not suffer from back-to-front reversal problems as do headphones. The reason for this is that loudspeakers can be physically placed in front of and behind the listener. Thus, if a certain sound source is to be played in front of or behind the listener, the sound source will physically emanate from the desired location; whereas with headphones, the sound source will only appear to emanate from the desired location.

Another advantage of loudspeakers is that a group can experience the added level of immersion, provided by sound, into a VE, as opposed to only one individual wearing headphones. For example, numerous people can be participating in the same virtual environment (i.e. fighting a battle in NPSNET). Furthermore, many groups of these people

40

will probably be located in the same location (i.e. a computer laboratory). Thus, placing loudspeakers in the laboratory will enable everyone in the room to experience the various sounds being generated in the virtual environment. Granted, the sounds will not be properly placed for all listeners, but still each listener in the group will be more immersed into the virtual environment as opposed to hearing no sounds at all.

Loudspeakers also have the advantage of being able to generate very low frequencies; whereas, "headphones do not allow listeners to feel low frequencies (below 150 Hz) via their body as a loudspeaker system...as real life does" [HENR91]. By using very low frequencies in the 4 Hz range, a greatly enhanced level of immersion is provided which is called *frequency injection* [ROES94].

2. Disadvantages

There are numerous disadvantages with generating 3D sound through free-field reproduction. One problem is that mismatched speakers (monitors) will severely degrade any attempt to spatialize the sound. Another problem is crosstalk which can occur when both ears receive the same sound from both loudspeakers [MART92]. Thus, left channel signals intended for the left ear are heard in the right ear and vice a versa. However, by proper use of transaural techniques, free-field crosstalk cancelation is possible [WENZ95]. Room acoustics also present numerous problems in trying to determine the best loudspeaker positions that produce the optimal listening environment.

Another problem with generating 3D sound through free-field reproduction is due the wave property of interference. This problem was touched upon earlier in the tuning fork experiment (see Figure 4). An extension of this experiment is to play a tone over loudspeakers in a large room. Then, as one walks around the room, one can also hear the tone appear to get louder and softer just like in the tuning fork experiment. These louder and softer spots in the room correlate to the nodes and antinodes of the tone as a result of the interference of the waves emanating from the speakers and from the various echoes of the room. As one can see, interference is one of the inherent problems of producing sounds in a free-field format. In trying to eliminate interference problems, one must ensure that the

listener is afforded the best possible listening area. This area is often called the *sweet spot*, the maximum convergence of all generated sound signals. As a result, if the listener is sitting in the sweet spot, the listener will be afforded the maximum potential listening environment. However, this sweet spot is static, so the listener's head must remain within the sweet spot in order to gain the benefits of the free-field sound system. This is perhaps the greatest disadvantage with using loudspeakers for use with VEs, for the size and position of the sweet spot of is relatively small and fixed. Thus, when a listener instinctively turns his head in an attempt to better reconcile a particular sound while in a VE, the listener will not gain any additional cues. This is because all 3D sound generated in a loudspeaker system is fixed according to the coordinate system of the loudspeakers as opposed to the real-life dynamic coordinate system of the moving head of the listener.

C.   CONCLUSION

It appears that headphone systems can better approximate actual real-time 3D sound through the use of individualized HRTFs when coupled with head-motion tracker systems. On the other hand, free-field systems, because of their openness to the environment, have greater inherent obstacles to over come. These inherent obstacles can be minimized by choosing properly matched quality loudspeakers that are very flat in magnitude and nearly linear in phase. As a result, crosstalk and other forms of unwanted interference are reduced. Additionally, because there are various applications of VEs, a headphone system might be more appropriate in one application; whereas a free-field system might be more applicable in another application. As such, since NPSNET was developed as a vehicle simulator, the orientation of one's immersion into the virtual world of NPSNET has traditionally been through some sort of vehicle (i.e. helicopter or tank). Thus, only vehicle actions were modeled and not those of individual head movements. So, the advantage of using headphone systems to isolate head movement was not needed. Therefore, the focus of this research is a continuation of presenting aural cues via free-field format oriented around vehicle actions use for use in NPSNET.

VI. NPSNET-3D SOUND SERVER

NPSNET-3D Sound Server (NPSNET-3DSS) is a MIDI-based free-field sound system consisting of "off-the-shelf" sound equipment and computer software which currently generates 2D aural cues for use in NPSNET, but is designed and capable of generating 3D aural cues. Its development is based on the previous NPSNET MIDI-based free-field sound systems and is the primary focus of this research.

A. GENERAL OVERVIEW

The approach taken in developing the NPSNET-3DSS was to build directly upon the previous NPSNET sound system: NPSNET-PAS [ROES94]. The basic concept was to enhance NPSNET-PAS from a 2D sound system to a 3D sound system. Accordingly, all 2D limiting factors had to be identified and improved. As a result, the hardware limitations of NPSNET-PAS sound generating equipment were identified and more capable "off-the-shelf" sound equipment was procured. Software limitations were also identified and a new algorithm was developed which properly distributes the total volume of a virtual sound source to a cube-like configuration of eight loudspeakers. It is this cube-like configuration of loudspeakers which forms the foundation for generating 3D sound. A second algorithm, based on the Precedence Effect, was also developed in an attempt to enhance one's ability to localize a sound source. This effort, however, proved unsuccessful. The final addition was adding synthetic reverberation through the use of digital signal processors to enhance perceptual distance of the generated 2D/3D aural cues. The resulting sound system of NPSNET-3DSS is similar to NPSNET-PAS but with some key changes. Figure 16 depicts the generalized structure of NPSNET-3DSS giving a good overview of the current system. It is important to understand this generalized view, for in the chapters to follow, many more details of this sound system will be presented.
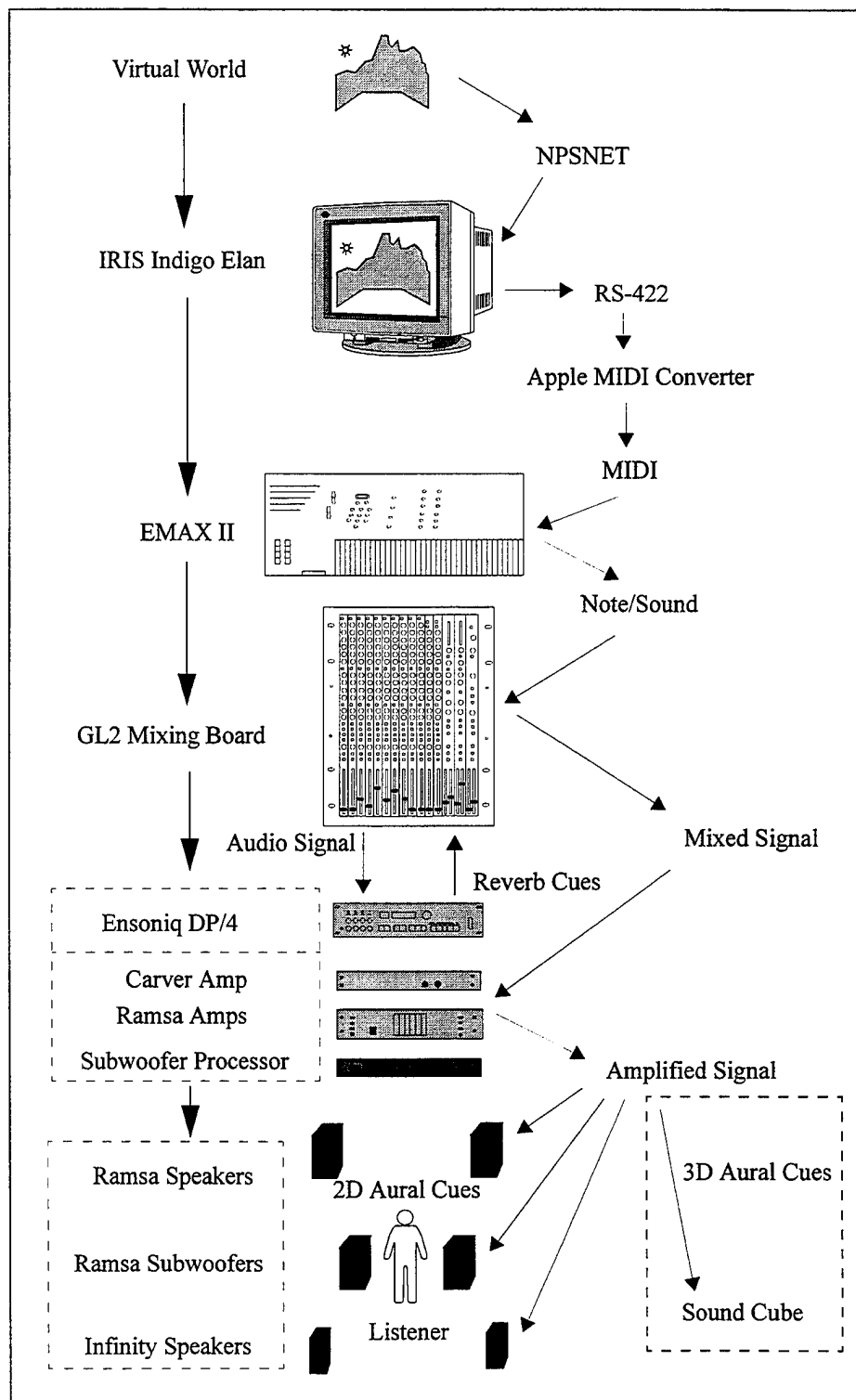
Figure 16: Overview of NPSNET-3DSS.

Virtual World

NPSNET

IRIS Indigo Elan

RS-422

Apple MIDI Converter

MIDI

EMAX II

Note/Sound

GL2 Mixing Board

Audio Signal

Mixed Signal

Reverb Cues

Ensoniq DP/4

Carver Amp

Ramsa Amps

Subwoofer Processor

Amplified Signal

Ramsa Speakers

2D Aural Cues

3D Aural Cues

Ramsa Subwoofers

Infinity Speakers

Listener

Sound Cube

B. SOUND CUBE CONCEPT

The sound cube concept is the heart of NPSNET-3DSS, for it is through this concept which enables the generation of 3D cues. The sound cube concept consists of a cube-like configuration of speakers and is depicted in Figure 17.
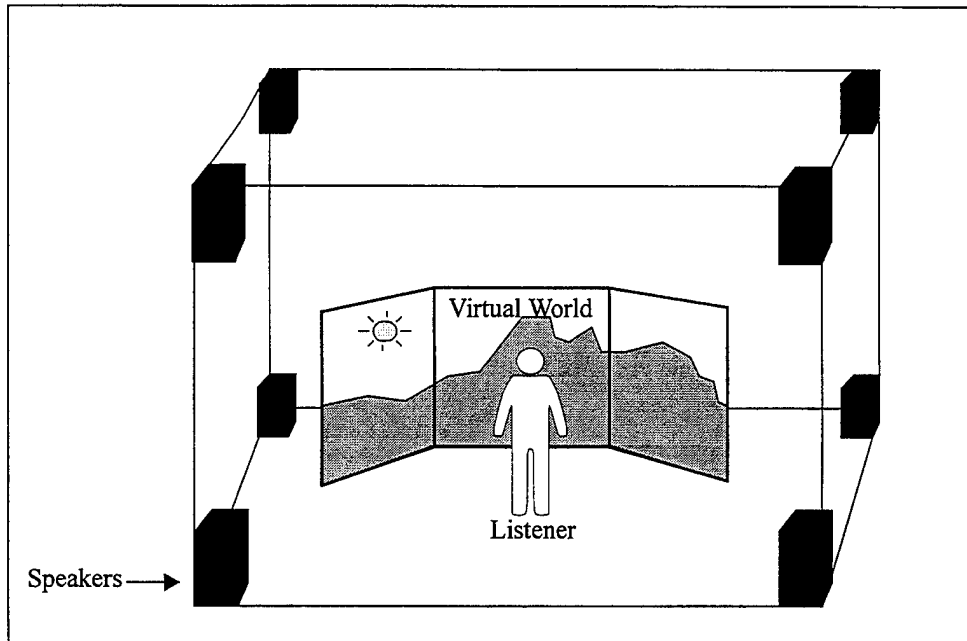
Figure 17: Sound Cube.

As seen in Figure 17, the active participant (listener) being immersed in our VE of NPSNET is located at the center of the cube of speakers. Specifically, it is the listener's head which must be located at the center of the sound cube, and not the center of mass of the listener. The reason for this placement is that the listener's head must be located completely within the sweet spot formed by all eight speakers. The front faces of all eight speakers point directly to this spot. As a result, this spot provides the only optimal position within the cube to uniformly hear sounds from all eight speakers. It should be noted that Figure 17 does not actually depict the correct angular displacement of the speakers. In order to ensure the widest possible sweet spot, the front faces of all the speakers would be perpendicular to the direction of the listener, which is the center of the cube. It is also

45

important that there are no obstacles between any of the speakers and the listener. Furthermore, there are numerous other concerns dealing with room acoustics which must be considered, but these concerns are beyond the scope of this research. The most important thing to gain from Figure 17 is a visualization of the sound cube concept.

1. The Problem

Given the cube configuration of speakers in Figure 17, the problem is to accurately represent the distance, direction, and volume of a sound source in the virtual world with respect to the listener by correctly distributing the total volume of this sound source among the eight speakers. This distribution of total volume among the various speakers is a form of sound localization. The sum of the volumes to be played from the individual speakers must be representative of the total volume of the original sound source. The end result is an apparent location of the sound source relative to the listener. It is this apparent sound source which provides an aural cue to the listener. Additionally, it is the combination of this aural cue with its associated visual cue which can dramatically increase one's immersion into not only NPSNET, but any VE. After finding an appropriate method to distribute the volume of the virtual sound source among the eight speakers, a generalized formula is needed which can be used for configurations of any numbers of speakers. The end result is a general mathematical sound model which can be used to localize sound via free-field format. This sound model is then capable of producing 2D or 3D localization cues depending on the numbers of speakers utilized. As such, in a quad configuration of four speakers, 2D cues are possible. In the cube-like configuration of eight speakers, 3D cues are possible.

2. Assumptions

Along with the problem to be solved, it is important to list the assumptions accepted before solving the problem. The assumptions are in the areas of sound source, listener, and the sound cube model (SCM) used.

a. Sound Source

In deriving the generalized SCM, it is assumed that only one sound source is to be played at any one time. This is of course not what happens in reality. In the real world many sounds are generated simultaneously. Accordingly, our sound model is not limited to playing only one sound source at any one time. The total number of possible sound sources which can be played by any sound system is a function of the capability of the particular sound generating equipment utilized. (In NPSNET-3DSS, the sound generating capability of the EMAX II permits sixteen simultaneous sounds. The EMAX II will be discussed in greater detail in a later chapter.) Nevertheless, a single sound source is used in the derivation of our sound model.

b. The Listener

A very critical assumption is that the listener's physical position in the sound cube is fixed relative to the speakers. As a result, the listener is always an equal distance from all eight speakers. Also, for the derivation of the sound model, it is assumed that the listener's heading and velocity are fixed. Again, this is not what happens in the real world, but it makes the derivation much easier.

c. The Sound Cube Model

We assume that the length of the sides of the sound cube model (SCM) are no shorter than the width of the listener's head. In other words, we assume that the listener's head fits completely within the SCM (see Figure 18). The reason for this assumption is that we are not allowing any sound sources to be played from within the listener's head. As a result, all sounds are externalized with respect to the listener's head. The length of the sides in the SCM is not to be confused with the actual length between the speakers in the sound cube configuration of Figure 17. The length between the speakers of the sound cube is dependent upon space available, room acoustics, the power/size of the speakers, and numerous others parameters. The distance used for NPSNET-3DSS is about eight feet.

47
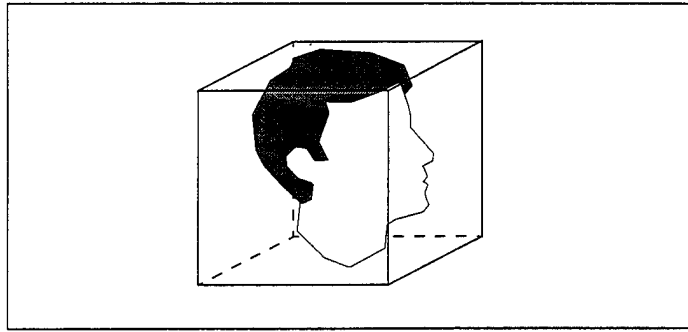
Figure 18: Sound Cube Model Related To Head Movement.

Another critical point to understand is that the speaker positions of the sound cube in Figure 17 correspond to the positioning of the vertices of the SCM. In other words, the sound cube is the actual physical implementation of the abstract mathematical SCM. Again, it is important to remember that these speaker positions are fixed with respect to the listener. Furthermore, there are two types of SCM's which can be implemented depending on how the listener interacts within the VE. If the listener is wearing a head mounted display (HMD) which corresponds to individual head movement, then the SCM must be related to the listener's head movement as depicted in Figure 18. If the listener is operating some sort of vehicle, and it is through this vehicle that the listener interacts within the VE, then the SCM must be related to vehicle movement as depicted in Figure 19.
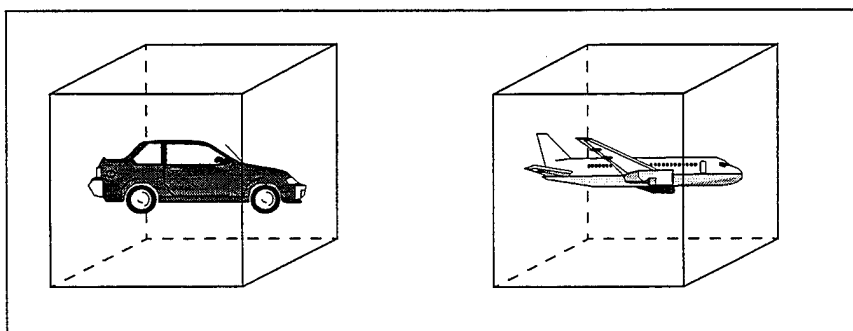
Figure 19: Sound Cube Model Related To Vehicle Movement.

NPSNET-3DSS is based on the SCM related to vehicle movement. Regardless of how the listener interacts within the VE, it is assumed that the listener's head will always be located within the dimensions of the sweet spot formed by the physical sound cube.

C.   REVIEW

Before continuing on to the next chapter, it is important to review the overall structure of NPSNET-3DSS and to be familiar with the listener's position within the sound cube. Furthermore, one must also have a good understanding of the SCM, for the next chapter presents the development of the generalized mathematical sound model which is used with NPSNET-3DSS.

VII. GENERALIZED 3D SOUND CUBE MODEL

The first step towards finding a generalized 3D Sound Cube Model (SCM) was to solve a 2D sound model. The concepts outlined in solving the 2D sound model form the foundation for understanding the 3D SCM.

A.   VOLUME

Determining the sound intensity/volume of a particular sound source within a VE is somewhat difficult and is still an active area of research. The work of Durand Begault, among others, is a standout in this area of research [BEGA91] [BEGA94]. One of Begault's basic ideas is that the volume of a sound source within a VE should not be based on traditional physically-based laws. For example, the physically-based formula for determining the intensity of a sound source, relative to distance, is the Inverse-Square Law (see Eq 12). In this formula, the intensity/volume, I, of a particular sound source, W, expressed in watts, is inversely proportional to the square of the radius/distance, r, from the listening point to the source. This correlates to a six decibel (dB) level reduction for each half-distance reduction [BEGA91].

$$I = W/4\pi r^2 \qquad\qquad \text{Eq 12}$$

Begault's work, however, suggests a that a more psychoacoustically-based formula is needed to calculate the volume of a sound source within a VE. In his work, Begault conducted several experiments in half-distance perception. In his experiments, a tone was played at some decibel level and was then increased and decreased. A test subject was then asked whether the perceived change in volume/intensity resulted in the perception that the sound had moved twice as far away or half the distance closer. Begault's work indicates that a reduction of more than six dB (from the Inverse-Square Law) is needed for each half-distance reduction. As a result, there is a much improved perception of half-distance. The exact decibel level of this reduction is not clear, for more experimentation is needed. However, the point is, the use of traditional physically-based laws does not work well for

51

determining the distance of a sound source within a VE. What is needed are psychoacoustically-based laws for determining the distance of a virtual sound source. Thus, based on Begault's findings, the following formula for volume was derived:

$$Volume = [1 - (\log_{10}(Distance/Half\_Dist)/\log_{10}(Max\_Range/Half\_Dist))] \times Total\_Volume$$

Eq 13

Distance is the length in meters from the source of a particular sound event to the listener. *Max_Range* comes from the maximum range at which a sound can be heard. *Half_Dist* is a constant used to represent the distance in which loudness decreases by some value more that 6 dB. *Total_Volume* is a constant representing the maximum volume of any sound that can be generated by our sound equipment. For example, the maximum volume for any sound using the MIDI protocol is 127 [INTE83]. This formula calculates the number of half-distances that the listener is away from the sound source. It then normalizes this number by the total number of half-distances within the *Max_Range*, using the *Half_Dist* number as the first half distance. The normalized number is now subtracted from *1* to give the appropriate percent volume that should be multiplied by the *Total_Volume*. In essence, the logarithmic nature of the intensity of sound is converted to a linear volume scale which can be easily implemented by most sound generating protocols (i.e. MIDI). [ROES94]

Substituting various values of *Max_Range* and *Half_Dist* allows one to control how far away a sound can be heard as well as it's drop-off rate. The current values utilized for *Max_Range* and *Half_Dist* are 12,700 meters and 25 meters respectively. These numbers were chosen mostly by trial and error through numerous demonstrations of NPSNET in an attempt to capture the appropriate perception of sound levels desired for use in NPSNET. A key factor in determining these values is the capability of the sound generating equipment. For example, if the volume of a particular sound source is calculated to have a MIDI note velocity of 40, the particular sound equipment utilized might not be able to generate a perceivable quality sound at this volume level. With this particular equipment, perhaps a higher range of MIDI note velocity is needed. So, not only is psychoacoustics important, but also the capability of the sound generating equipment. Better capable

equipment will result in more realistic sounds due to their increased dynamic range. But no matter what formulas or equipment is used, the most important factor is the listener's perception of the generated sounds. The choice of *Max_Range* and *Half_Dist* is still an ongoing area of research.

B.   SPEED OF SOUND

Before any sound can be distributed among the various speakers, we must know when to play this sound. The time to play a sound source within our VE corresponds to the distance between the listener and the sound source. The time it takes this sound source to travel to the listener is based on the speed of sound. Thus, when a sound event occurs in our VE, we simply measure the distance between the listener and the sound source and divide this distance by the speed of sound. The result gives us the appropriate amount of delay time to compensate for the speed of sound. The speed of sound used in this research is normalized to sea level at 70 degrees Fahrenheit, in air, at 335.28 meters per second. There are numerous other parameters besides the speed of sound which need to be taken into consideration in determining when a sound source is to be played. However, these other parameters are beyond the scope of this paper, and many are still active areas of research.

C.   2D SOUND MODEL

Given Eq 13, which calculates the total volume of a sound source within our VE, and using the speed of sound to determine when to play this sound, we can now distribute this volume of sound among the speakers. For the development of the 2D sound model, we use a sound system consisting of four speakers. Figure 20 represents how the 2D sound model corresponds to these speaker locations.

The amount of sound to be distributed among the various speakers correlates to a percentage of the total possible volume of the sound source. To calculate the percentage of volume to be played at each speaker, we cast out two different types of vectors from the listener as seen in Figure 20. The first type of vector is from the listener to each speaker.

Figure 20: 2D Sound Model.

S = Sound Source
L = Listener
A,B,C,D = Correlate to Speaker Positions
θ_{SA} = The Smaller Angle BetweenVectors LS & LA
$\theta_{AB} = \theta_{BC} = \theta_{CD} = \theta_{AD} = 90°$
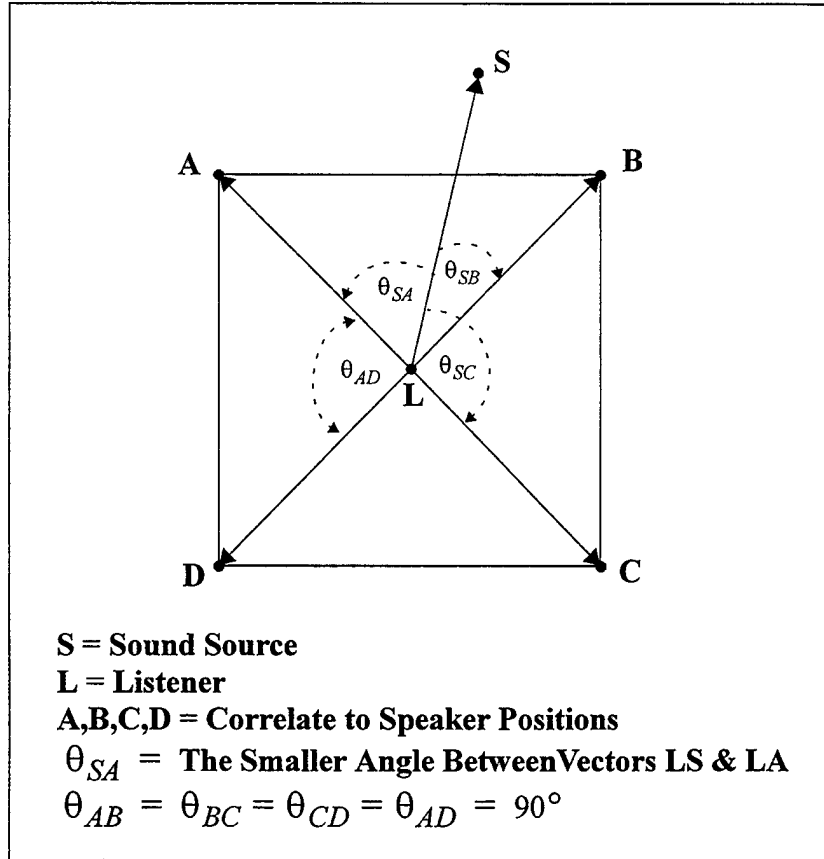
There are four of these vectors: $\overrightarrow{LA}$, $\overrightarrow{LB}$, $\overrightarrow{LC}$, and $\overrightarrow{LD}$. The second type of vector is from the listener to the source. There is only one of this type vector: $\overrightarrow{LS}$. Using the dot product, we can determine the angles between vectors $\overrightarrow{LS}$ and $\overrightarrow{LA}$, $\overrightarrow{LB}$, $\overrightarrow{LC}$, and $\overrightarrow{LD}$. We will call these angles θ_1, θ_2, θ_3, and θ_4 respectively. For example, in Figure 20, $\theta_1 = \theta_{SA}$, and $\theta_2 = \theta_{SB}$. Observe that the angle formed between $\overrightarrow{LA}$ and $\overrightarrow{LB}$, $\overrightarrow{LB}$ and $\overrightarrow{LC}$, etc. is 90 degrees. The importance of this angle is described later.

In looking at Figure 20, we see that the source S is located somewhere between A and B. Remember that A and B correspond to speaker locations. Thus, the speakers that should play the sound source should be speakers A and B. Furthermore, A and B should be the only speakers generating sounds and not speakers C or D. It should be fairly intuitive

that speakers *A* and *B* are the only speakers which need to play the sound source for they are the closest to the sound source. In this case, if any portion of the sound were to emanate from any other speaker, the proper localization of sound relative to the listener would be lost.

Observe that the angles formed between vectors $\overrightarrow{LS}$ & $\overrightarrow{LA}$, and $\overrightarrow{LS}$ & $\overrightarrow{LB}$ are less than 90 degrees. And, the angles formed between vectors $\overrightarrow{LS}$ & $\overrightarrow{LC}$, and $\overrightarrow{LS}$ & $\overrightarrow{LD}$ are greater than 90 degrees. The importance of the 90 degree angle is now apparent. If the angle formed between the sound source and the speaker, relative to the listener, is greater than 90 degrees, we discard the possibility of playing any sounds from the associated speakers. If, on the other hand, the angle formed between the sound source and the speaker, relative to the listener, is less than 90 degrees, the associated speakers are the only ones with the possibility of playing any sounds. Thus, a maximum of two speakers is all that can be played for each sound source. The sound model is also optimized for speed, for it discards half of the possible speaker combinations before calculating the percentage of volume to be played at each speaker. This optimization for speed helps to ensure that all sounds are generated in real-time -- a vital requirement for any VE. The method to calculate this percentage is described later.

Another factor that has to be considered is when the sound source is in close proximity to one of the lines formed by the listener/speaker vectors. For example, if the sound source is located at a position corresponding to the exact direction of one of the speakers, then it would only be necessary to play the sound at that speaker and no other speaker. Thus, in the sound model we also test for how close a sound source is to the direction of any one speaker. If the sound source is within three degrees of any one speaker, relative to the angle formed between the listener and the speaker, then only that speaker will play the sound. Again, because we want to optimize the sound model for speed, this close proximity check will eliminate the other speakers before calculating the percentage of volume to be played. The decision to use three degrees was chosen somewhat arbitrarily. The number of degrees to use or even the idea of using this close proximity check is an area

of ongoing research. Nevertheless, three degrees seems to works very well with this sound model.

Now that we have identified which speakers to play, we need to properly distribute the total volume of the sound source among these speakers. The following formula has been derived to distribute this total volume:

$$V_i = V_{total}\left[1 - (n-1)\frac{\theta_i}{sum}\right]$$

Eq 14

V_i is the volume to be played at each respective speaker, where $i = 1$ corresponds to speaker A, and $i = 2$ corresponds to speaker B, etc. V_{total} is the total volume of the sound source calculated from Eq 13. θ_i, as mentioned above, corresponds to the angles formed between vectors $\overrightarrow{LS}$ and $\overrightarrow{LA}$, $\overrightarrow{LB}$, $\overrightarrow{LC}$, and $\overrightarrow{LD}$. For example, as shown in Figure 20, $\theta_1 = \theta_{SA}$ and $\theta_2 = \theta_{SB}$. *sum* is the summation of all angles θ_i, where θ_i is less than 90 degrees. n is the number of angles θ_i, in which θ_i is less than 90 degrees. In the 2D sound model, this number n has a maximum value of 2. Thus, for any given n, and θ_i less than 90 degrees, the *sum* must be constrained as follows:

$$sum = \sum_{i=1}^{n} \theta_i$$

Eq 15

Also, since the formula in Eq 14 is normalized, the total volume must also be constrained as follows:

$$V_{total} = \sum_{i=1}^{n} V_i$$

Eq 16

We now have all that is needed to properly distribute the total volume of a sound source among the various speakers in a 2D sound system. Notice that Eq 14 indicates an inverse proportional relationship between θ_i and V_i. Thus, if θ_i is small, then V_i is large and visa versa. This inverse proportional relationship between θ_i and V_i is the foundation of the general nature of this sound model.

D.   3D SOUND CUBE MODEL

Given the 2D sound model, we can easily generalize this 2D model to the 3D sound cube model (SCM). We use the same formula for calculating the volume of the sound source within our VE (see Eq 13). We also continue to use the speed of sound to determine when to play this sound source. All we need to do is recalculate how to distribute the total volume of the sound source from among four speakers to eight speakers. The new 3D SCM can be seen in Figure 21. The listener is now located in the center of a cube. Like the 2D

S = Sound Source
L = Listener
A,B,C,D,E,F,G,H = Correlate to Speaker Positions
θ_{SA} = The Smaller Angle BetweenVectors LS & LA
$\theta_{AB} = \theta_{BC} = \theta_{CD} = \theta_{AD} = 70.5°$
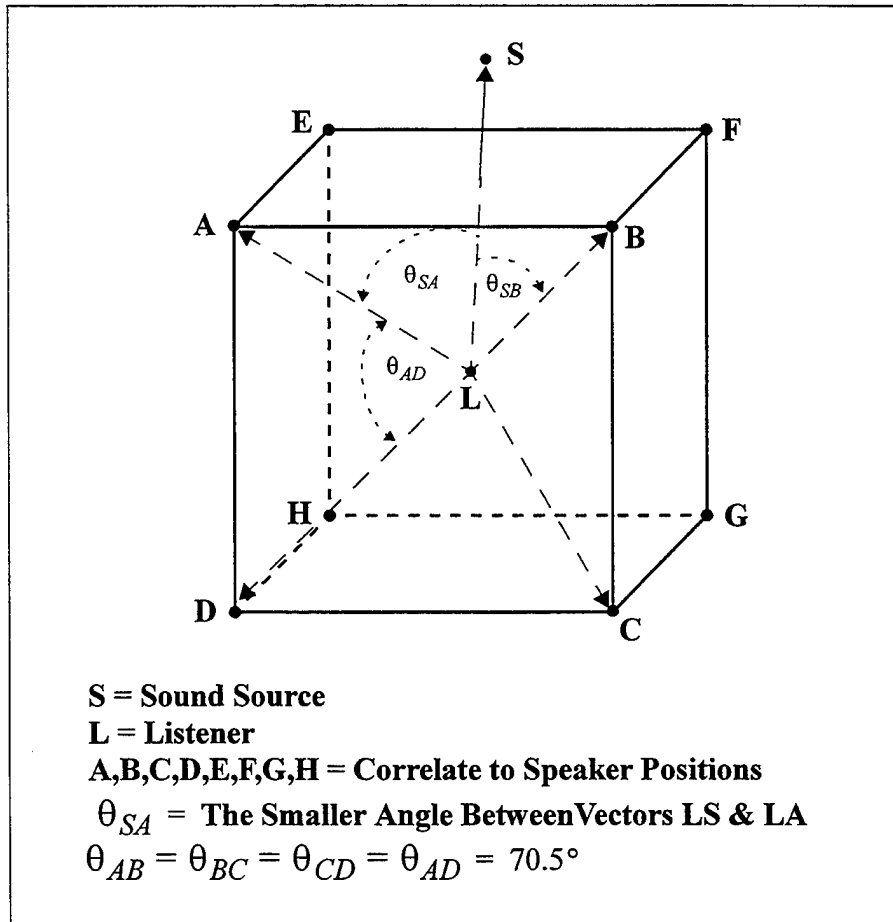
Figure 21: 3D Sound Cube Model.

model, the amount of sound to be distributed among the various speakers still correlates to a percentage of the total possible volume of the sound source. The calculation of this percentage is the same as in the 2D model except that now we have twice the number of

speaker positions. To calculate the percentage of volume to be played at each speaker, we cast out two different types of vectors from the listener as seen in Figure 21. The first type of vector is from the listener to each speaker. There are now eight of these vectors: $\overrightarrow{LA}$, $\overrightarrow{LB}$, ... ,$\overrightarrow{LH}$. The second type of vector is the source vector: $\overrightarrow{LS}$. We use the dot product to determine the angles between vectors $\overrightarrow{LS}$ and $\overrightarrow{LA}$, $\overrightarrow{LB}$, ... ,$\overrightarrow{LH}$. Again, we will call these angles θ_1, θ_2, ... ,θ_8. Observe that now the angle formed between $\overrightarrow{LA}$ and $\overrightarrow{LB}$, $\overrightarrow{LB}$ and $\overrightarrow{LC}$, etc. is approximately 70.5 degrees. So now, when looking at Figure 21, we can see that the source S is located somewhere between A, B, E, and F. Thus, the only speakers that should play the sound source should be speakers A, B, E, and F. If any portion of the sound were to emanate from any other speaker, the proper localization of sound relative to the listener would be lost.

Observe that the angles formed between vectors $\overrightarrow{LS}$ & $\overrightarrow{LA}$, $\overrightarrow{LS}$ & $\overrightarrow{LB}$, etc. must now be less than 70.5 degrees. And, the angles formed between vectors $\overrightarrow{LS}$ & $\overrightarrow{LC}$, $\overrightarrow{LS}$ & $\overrightarrow{LD}$, etc. must be greater than 70.5 degrees. If the angle formed between the sound source and the speaker, relative to the listener, is greater than 70.5 degrees, we discard the possibility of playing any sounds from the associated speakers. If, on the other hand, the angle formed between the sound source and the speaker, relative to the listener, is less than 70.5 degrees, the associated speakers are the only speakers to be played. Thus, with this 3D SCM a maximum of four speakers is all that can be played for each sound source. Again, our sound model is optimized for speed, for it discards half of the possible speaker combinations before calculating the percentage of volume to be played at each speaker.

For the case when the sound source is in close proximity to one of the lines formed by the listener/speaker vectors, we use the same methodology as in the 2D model. If the sound source is within three degrees of any one speaker, relative to the angle formed between the listener and the speaker, then only that speaker will play the sound. Again, because we want to optimize our sound model for speed, this close proximity check will

eliminate the other seven speakers before calculating the percentage of volume to be played. As before, this is still an ongoing area of research.

Now that we have found which speakers to play, we need to properly distribute the total volume of the sound source among these speakers. Because of the general nature of our sound model, we can use the same formula as in the 2D model as shown before in Eq 14 as follows:

$$V_i = V_{total}\left[1 - (n-1)\frac{\theta_i}{sum}\right]$$

V_i is the volume to be played at each respective speaker, where $i = 1$ corresponds to speaker A, and $i = 2$ corresponds to speaker B, etc. V_{total} is the total volume of the sound source calculated from Eq 13. θ_i corresponds to the angles formed between vectors $\overrightarrow{LS}$ and $\overrightarrow{LA}$, $\overrightarrow{LB}$, ... $\overrightarrow{LH}$. sum is the summation of all angles θ_i, where θ_i is less than 70.5 degrees. n is the number of angles θ_i, in which θ_i is less than 70.5 degrees. In our 3D SCM, this number n has a maximum value of four.

Again, for any given n, and θ_i less than 70.5 degrees, the sum must be constrained as shown previously in Eq 15 in the 2D model as follows:

$$sum = \sum_{i=1}^{n} \theta_i$$

Furthermore, as in the 2D model, since the formula in Eq 14 is normalized, the total volume must also be constrained as previously shown in Eq 16 as follows:

$$V_{total} = \sum_{i=1}^{n} V_i$$

Again, we now have all that is needed to properly distribute the total volume of sound among the various speakers in a 3D sound system. As can be seen, the inverse proportional relationship between θ_i and V_i is still valid in our 3D SCM.

VIII. PRECEDENCE EFFECT SOUND MODEL

As the name implies, this sound model is based on the Precedence Effect (PE) (see The Precedence Effect on page 24). As such, to base a sound model on the PE, an entirely different approach has to be taken as opposed to that used in the development of the 3D sound cube model (SCM). There are, however, a few similarities with the SCM as follows:

- Both use the same sound cube (SC) speaker configuration depicted in Figure 17.
- Both use the same method to calculate the delay time to play a sound source based on the speed of sound.
- Both use the same psychoacoustically-based formula to calculate the volume of a virtual sound source as shown in Eq 13.

Besides these similarities, the PE sound model is radically different from the SCM.

In the SCM we were only interested in the location of the sound source; whereas in the PE sound model we are interested not only in location of the sound source, but also in the resulting sound waves (see Wave Properties of Sound on page 17). Thus, by further modeling the generated sound waves of the sound source, the PE sound model attempts to better emulate how we hear sounds in the real world. In looking at Figure 22, we see the sound source and its resulting sound waves which travel at the speed of sound. Although not depicted as such, these sound waves should be thought of as three-dimensional spheres emanating from the sound source S. The basic idea of the PE sound model is to play the appropriate volume of the sound source upon the intersection of the sound wave with the speaker position. For example, when the sound wave reaches the position which correlates to speaker A, we play the volume of the sound source at speaker A. When the sound wave reaches the position which correlates to speaker B, we play the volume of the sound source at speaker B, etc. Unlike the SCM which plays the sound at a maximum of four speakers, the PE sound model always plays the sound at all eight speakers of the SC as depicted in Figure 17. The final result is an attempt to emulate the sound wave as it passes through the listener. In looking at Figure 22, if we imagine that the sound source S is emanating at a
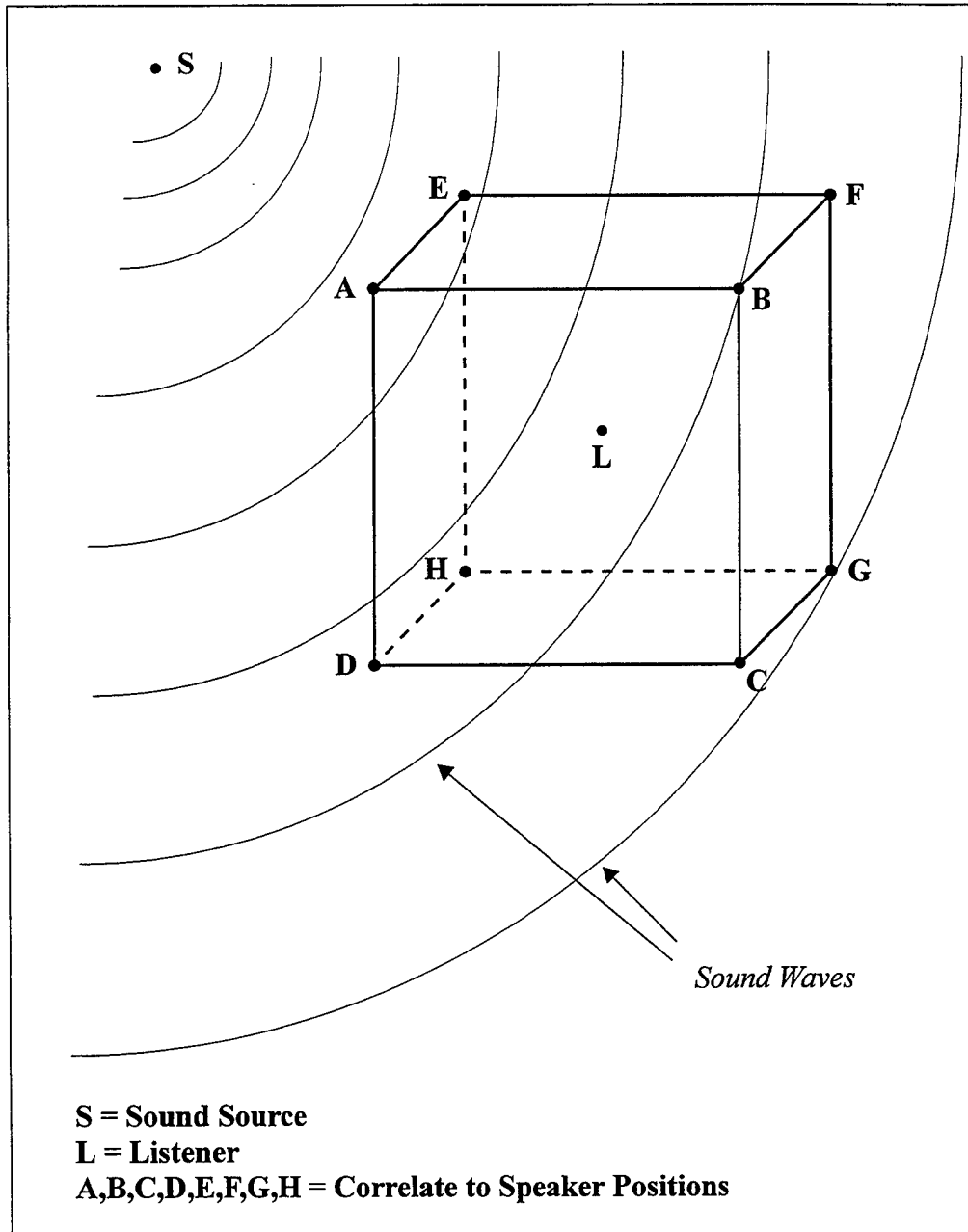
Figure 22: Precedence Effect Sound Model.

distance forward of the listener L (this would be somewhere in the direction towards the inside of this page), then the speaker which correlated to position E would be the first to play the sound. The other speakers would then play the sound according to when the sound wave intersected their corresponding positions. Thus, based on the PE, since the listener heard the sound first from position E, the listener would perceive that the sound was located

in the direction of position *E*. As a result, the listener can correctly localize the sound source. However, since the PE is only effective within the first 30 ms of hearing a sound source [EVER91b], the difference in time when all eight speakers play the sound cannot exceed 30 ms. If this time constraint is exceeded, the listener will no longer perceive a single sound source, but instead multiple sound sources making localization of the original intended sound source impossible.

> With the case of two impulses [sounds] spaced closely in time, the separation of these two impulses determines a wide range of perceptual effects. Certainly if the two pulses are more than 30 to 50 milliseconds apart, they will be heard as two separate and distinct pulses. [MOOR79]

Therefore, for this PE sound model to be effective, its corresponding sound system must be able to generate sounds to all eight speakers within 30 ms.

IX. SYNTHETIC REVERBERATION

Just as the SCM and the PE sound model attempt to generate appropriate cues to aid in localizing a sound source for use in a VE, so does synthetic reverberation (SR) attempt to help localize a sound source. Both the SCM and the PE sound model are used to generate the intensity (volume) of the sound source which is perhaps the most important cue in sound localization. SR, on the other hand, attempts to add the lesser important localization cue of reverberation to the sound source (see Sound Localization on page 19 for other localization cues). However, the extent of the importance of reverberation in sound localization is an active area of research. It is important to note that SR can only be used in conjunction with a predetermined intensity level of some sound source. In the case of this research effort, the intensity can be derived from either the SCM or the PE sound model, but any method for determining intensity can be utilized. The basic idea is that SR means nothing without an associated intensity.

A.   BACKGROUND

The use of SR is based on the fact that reverberation adds a very important physical and psychoacoustic quality to sound. The Journal of the Acoustical Society America (JASA) defines sound as having three qualities: 1) pitch, 2) intensity, and 3) tamber (also called *timbre* which refers to anything not in pitch or intensity). As such, reverberation falls into the category of tamber, and therefore helps to define the overall characteristic of the sound. To gain a better appreciation for the defining characteristic of reverberation, we can look at the makeup of a tone. There are three parts to a tone: 1) attack, 2) steady state, and 3) decay. In looking at Figure 23, we can see the temporal displacement of these three parts. The last part of the tone is decay which is mostly a function of reverberation. By using different amounts of reverberation to produce varying lengths of decay, we can produce different sounding tones. This is the whole idea behind using SR, in that we can recreate a particular characteristic of sound by manipulating the tamber of the sound through
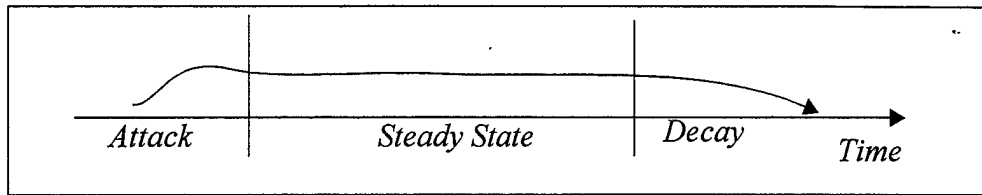
Figure 23: Three Parts to a Tone.

judicious choice of reverberation. Various amounts and types of reverberation can be produced synthetically through the use of digital signal processors (DSPs).

B.   PREVIOUS APPLICATIONS

The study of reverberation dates back to 1900 when W. Sabine examined room reverberations [SABI72]. The first published computer simulations of room reverberation was done by M. Schroeder in 1961/1962 [SCHR61] [SCHR62]. Schroeder's work provided the foundation for artificially generating reverberation. The mechanism through which this artificial reverberation was generated consisted of a unit reverberator using an all-pass filter or a comb filter. The unit reverberator is the oldest ancestor of the DSP.

1. Moorer/IRCAM

In 1978, J. Moorer from the Institut de Recherche et Coordination Acoustique/ Musique (IRCAM: the Institute of Research and Coordination of Acoustics and Music) showed that the then existing reverberation techniques were not accurate. One of Moorer's conclusions was that "all the geometric simulations of concert hall acoustics that have been done to date result in a simulated room reverberation that does not sound at all like real rooms" [MOOR79]. Furthermore, he found "a much larger number of non-useful unit [reverberation] generators than useful new unit generators" [MOOR79].

2. Chowning/CCRMA

In 1982, J. Chowning and C. Sheeline from the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University conducted experiments of auditory distance perception using SR [CHOW82].

66

The primary objective of this project was the development of a practical method for generating perceptual conditions of a realistic and room-like nature, for the purpose of testing the ability of humans to judge the source distance of sound. [CHOW82]

In their experiment, Chowning and Sheeline recorded a trumpet sound in a dead room. This recorded sound was then played back and recorded in auditoria of various sizes on Stanford's campus. Chowning and Sheeline then used basically the same reverberation algorithm developed earlier by Moorer to recreate the ambient conditions of different auditoria on Stanford's campus. A test subject was then asked to listen to various pairs of rooms chosen from both the actual recorded sounds and the synthetically recorded sounds. One of the general conclusions of this experiment was that "the most salient characteristic for all listeners, when asked to differentiate among listening spaces, is that of reverberation time" [CHOW82].

3. Begault/NASA-Ames Research Center

In 1991, D. Begault from NASA-Ames Research Center at Moffett Field conducted an experiment on the perceptual effects of using SR [BEGA92]. In this experiment, five test subjects were presented a segment of speech via headphones. The speech segment was processed using nonindividualized head-related transfer functions (HRTF). (See Head-Related Transfer Function on page 36.) Furthermore, the speech stimuli was processed both with and without spatial reverberation generated via a DSP. The test subjects were presented with the speech stimuli and were then asked to estimate its azimuth and elevation. The results of this study showed that when SR was added to the speech stimuli, the test subjects experienced a more realistic externalization of the sound. However, the added SR caused an increase in azimuth and elevation localization errors. In terms of distance perception, "All subjects made relative increases in their distance judgements when reverberation was added to the stimuli" [BEGA92].

4. Brungart/Wright State University

In 1993, D. Brungart from Wright State University conducted an experiment on distance simulation in virtual acoustic displays [BRUN93]. In this experiment six subjects were asked to make distance judgements of white noise presented via free-field and headphone delivery systems. Half of the tests included only the intensity of the white noise, and the other half included the intensity along with SR generated by a DSP. The white noise simulated distances ranging from two to nineteen feet from the listener. The results of this experiment showed that the test subjects were able to correctly identify the distances of the white noise up to ten feet via free-field format. However, when the SR was added to the white noise in the free-field format, the judgements of the test subjects were one to two feet longer for distances beyond ten feet. When the test subjects repeated the experiment via headphones including only the intensity of the white noise, they overestimated distances less than ten feet and underestimated distances beyond ten feet. Furthermore, when SR was added to the white noise, the results were virtually identical to the white noise only case via headphones. Thus, the results of using SR via free-field format is inconsistent with the results via headphone systems. [BRUN93]

C. APPLICATION IN VIRTUAL ENVIRONMENTS

This research focuses on the use of SR in virtual environments (VEs) to recreate ambient environments and to increase distance perception.

1. Ambient Environment

Just as SR can be used to help recreate an acoustic ambient condition (i.e. a room, concert hall, auditorium, etc.), so can SR be used to help recreate the ambient environment within a virtual world. As in the previous uses of SR, DSPs can be used to produce the required SR. However, the emphasis on using SR has been traditionally centered on how to reproduce various *inside* conditions such as a small or large room. As such, most commercial off-the-shelf DSPs include reverberation algorithms reproducing these inside

conditions of a small or large room as opposed to outside conditions. One reason for the bias towards producing these inside condition reverberation algorithms is because the earlier applications of reverberation centered primarily on recreating musical environments such as concert halls. Another reason is that inside conditions are simpler and more standardized. For example, a room typically has a floor, ceiling and four walls having fairly common dimensions. The outdoors, though, has no typical reflection surfaces and no common dimensions, but "can be approximated by assuming a single floor reflection" [BRUN93]. Thus, it is possible to use commercial off-the-shelf DSPs to recreate both inside and outside ambient environments. The question remains, how can DSPs be utilized to recreate ambient conditions for use in a VE?

In line with the MIDI-based sound system of NPSNET, this research proposes using a DSP with Musical Instrument Digital Interface (MIDI) capabilities. The basic idea is to send a MIDI command to the DSP, which would in turn select a certain reverberation algorithm. The particular reverberation algorithm selected is based on the virtual world coordinates of the immersed user. For example, when an NPSNET player enters a building, a MIDI command is sent to the DSP which would change the reverberation algorithm to that of possibly a small or large room. Today's commercial off-the-shelf DSPs are very capable having preprogrammed reverberation algorithms of many types including small rooms, large rooms, concert halls, etc. However, if these factory preset algorithms are not suitable, most common DSPs allow for changing various parameters of these algorithms to produce a customized desired effect. Thus, within a VE, bounding volumes of particular areas (such as buildings, valleys, caves, etc.) can be associated with any desired reverberation effect ultimately producing a more realistic acoustic mapping of the VE. The only requirement of the DSPs is to be able to change reverberation algorithms in real-time with no perceivable loss of sound.

2. Distance Perception

As indicated in the earlier studies, adding SR along with the appropriate intensity increases the perception of distance of a sound source. Again, in line with the MIDI-based sound system of NPSNET, the basic idea is to send a MIDI command to the DSP, which in turn selects a certain reverberation algorithm. In this case, the particular reverberation algorithm selected is based on the distance between the immersed listener and the sound event. For example, an NPSNET player sees and hears an explosion of some type at approximately 100 meters away. A MIDI command is then sent to the DSP which in turn selects a reverberation algorithm producing some amount of reverberation/decay of the explosion. Now, the same NPSNET player sees and hears an explosion of some type at approximately 500 meters away. A MIDI command is again sent to the DSP, but this time the reverberation algorithm selected produces a relatively greater amount of reverberation/ decay of the explosion. Thus, an algorithm based on the distance from the listener to the sound source can be applied to any sound event in the VE ultimately selecting appropriate reverberation/decay for increased distance perception. Again, the only requirement of the DSPs is to be able to change reverberation algorithms in real-time with no perceivable loss of sound.

X. SOFTWARE AND HARDWARE FUNCTIONALITY

This chapter discusses the main software and hardware functionality of the NPSNET-3DSS. Specifically, the software functionality discusses how sound events in the VE of NPSNET are identified and processed by the NPSNET-3DSS. The hardware functionality describes the hardware interface between NPSNET and the NPSNET-3DSS along with a description of the configuration and use of NPSNET-3DSS sound equipment.

A. SOFTWARE FUNCTIONALITY

Except for some minor changes, the overall software design and functionality of NPSNET-3DSS is virtually identical to that of its predecessor, NPSNET-PAS. For a full description of the software functionality see Roesli's master's thesis [ROES94]. However, a brief overview follows.

The primary purpose of the main function is to monitor the Distributed Interactive Simulation (DIS) packets being generated in the network for which NPSNET is operating (see Figure 24). From these DIS packets, if there is Protocol Data Unit (PDU), the main function will then process the PDU. There are currently three PDUs which have an associated sound event: 1) Entity State PDU, 2) Fire PDU, and 3) Detonation PDU. The Entity State PDU is used to process the host vehicle sound actuation and acceleration. The host refers to the particular machine (i.e. Meatloaf, Elvis, Gravy3, etc.) for which the aural cues are being generated. The Entity State PDU is processed by the function *process_entityPDU*. The Fire PDU is used to process the firing of some sort of weapon belonging to the host or any entity capable of firing a weapon. The Fire PDU is processed by the function *process_firePDU*. The Detonation PDU is used to process all weapon detonations/explosions and is processed by the *process_detonationPDU*. After all PDUs have been processed, the *process_state* function updates and manages several control functions concerning the state of the host and NPSNET-3DSS functions. After all state functions have been processed, a dead reckoning algorithm is used to update the host's

Main Function Initialization
- Process Command Line
- Set Up Window
- Read Configuration File
- Get Host Data
- Load Environmental Data
- Open Network

Process PDUs
- Process Fire PDUs
- Process Detonate PDUs
- Process Entity State PDUs

Begin Main Loop

If we have a PDU

All PDUs processed or 5 have been processed

No PDUs to process

Update Event List

If no events

Process State

If event occurs

Trigger 3D Sound

Dead-Reckon Host

Draw Graphic Display

Process Keyboard Input

Process Environmentals

If exit key pressed

End Program

Figure 24: NPSNET-3DSS Program Flow. From [ROES94].

72

position in the virtual world. Next, the *update_event_list* function updates all possible sound events based on the speed of sound ultimately determining when to play a sound event. Currently, if a sound event is beyond 12700 meters, it is deleted from the list. When it is time to play a sound, the function *trigger_3D_sound* generates the appropriate MIDI commands to physically play the particular sound. This function is the heart of the NPSNET-3DSS, for it generates the 2D/3D spatialized localization aural cues to the host NPSNET player. This function will be described in much greater detail in Chapter XI. IMPLEMENTATION AND ANALYSIS. Next, the 2D graphic display of the host and all sound events are updated and redrawn by the *update_window* function. In looking at Figure 25, the 2D graphic display is depicted where *F* represents a Fire PDU and *D* represents a Detonation PDU. Associated with each sound event is an increasing circle representing the
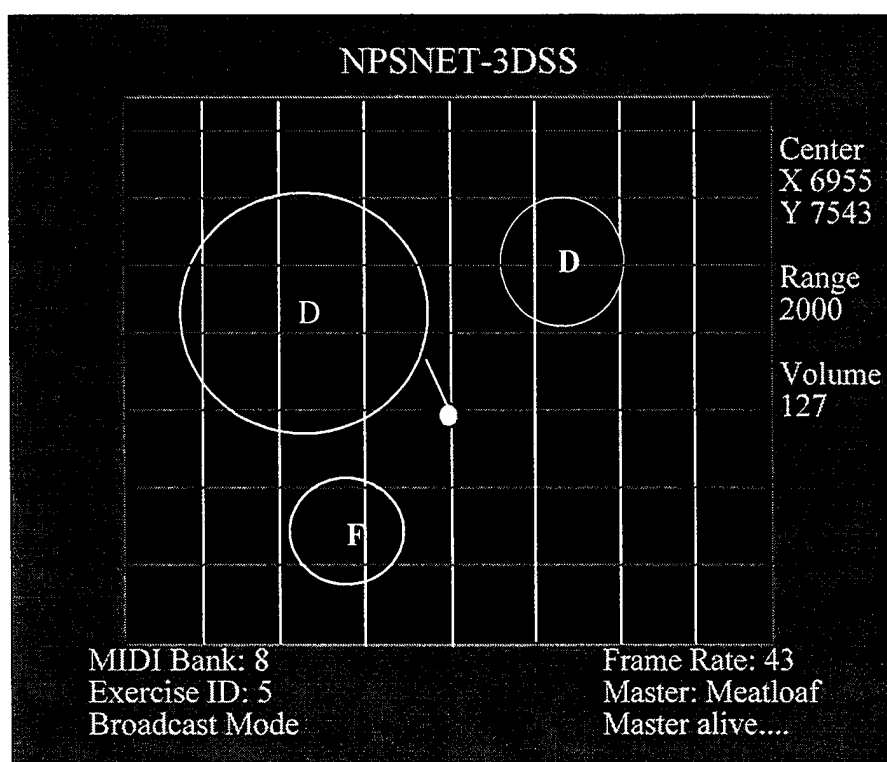
Figure 25: NPSNET-3DSS 2D Graphic Display. After [ROES94].

sound wave of the sound event traveling at the speed of sound. At this point, any environmentally related cues based on the host's position is the virtual world are processed

by the function *process_environmentals*. It is in this function where various reverberation algorithms can be sent to a DSP to recreate the ambient conditions of a building, cave, valley, etc. The last function called is *process_keyboard*, which manages possible input from the keyboard such as the *escape* key which will terminate the NPSNET-3DSS program. All the aforementioned functions reside in the main program loop.

B. HARDWARE FUNCTIONALITY

The hardware functionality of NPSNET-3DSS has two aspects: 1) partial sound cube (SC) implementation and 2) full SC implementation. The following discussion describes these two aspects along with a description of the overall hardware flow of NPSNET-3DSS.

1. Partial Sound Cube Implementation

Currently the hardware for NPSNET-3DSS consists of the following:
- One (1) IRIS Indigo Elan.
- One (1) Apple MIDI Interface Converter.
- One (1) EMAX II Digital Audio Sampler/Sequencer.
- One (1) GL2 Allen and Heath Mixing Board.
- Two (2) Ensoniq DP/4 Digital Signal Processors.
- One (1) Ramsa Subwoofer Processor.
- One (1) Carver Power Amplifier.
- Two (2) Ramsa Power Amplifiers.
- One set (2 total) of Ramsa Subwoofers.
- One set (2 total) of Infinity Speakers.
- One set (2 total) of Ramsa Studio Monitors.

Along with this hardware are numerous types of cables for routing audio and MIDI signals. The specific wiring diagrams representing the actual interface connections for all the various pieces of hardware of the current NPSNET-3DSS are depicted in APPENDIX C: HARDWARE WIRING DIAGRAMS. The basic hardware configuration of the current

NPSNET-3DSS is depicted in Figure 26. This is only a temporary configuration for it lacks the additional amplifiers and speakers needed to create the sound cube (SC) as depicted earlier in Figure 17. Until receipt of this additional equipment, NPSNET-3DSS is currently implemented using the same speaker placement of NPSNET-PAS as depicted earlier at the bottom of Figure 3. As a result, this system can only produce 2D aural cues. Nevertheless, the underlying foundation of this current system is still centered around using the equipment required for the SC. But, since this current system has only four speakers, as opposed to eight speakers needed for the SC, this system collapses the SC from three dimensions to two dimensions. In looking at Figure 26, the eight audio signals generated by the EMAX II (which would be sent to the eight speakers of the SC) are sent to only four speakers. In essence the 3D cube is squashed into a 2D square representing the speaker placement of NPSNET-PAS. Therefore, this current system is fully capable of generating the required 3D spatialized aural cues but simply lacks the additional amplifiers and speakers needed for the SC.

2. Full Sound Cube Implementation

To fully implement the SC, the hardware for NPSNET-3DSS must consist of the following:

- One (1) IRIS Indigo Elan.
- One (1) Apple MIDI Interface Converter.
- One (1) EMAX II Digital Audio Sampler/Sequencer.
- One (1) GL2 Allen and Heath Mixing Board.
- Two (2) Ensoniq DP/4 Digital Signal Processors.
- One (1) Ramsa Subwoofer Processor.
- Five (5) Ramsa Power Amplifiers.
- One set (2 total) of Ramsa Subwoofers.
- Four sets (8 total) of Ramsa Studio Monitors.
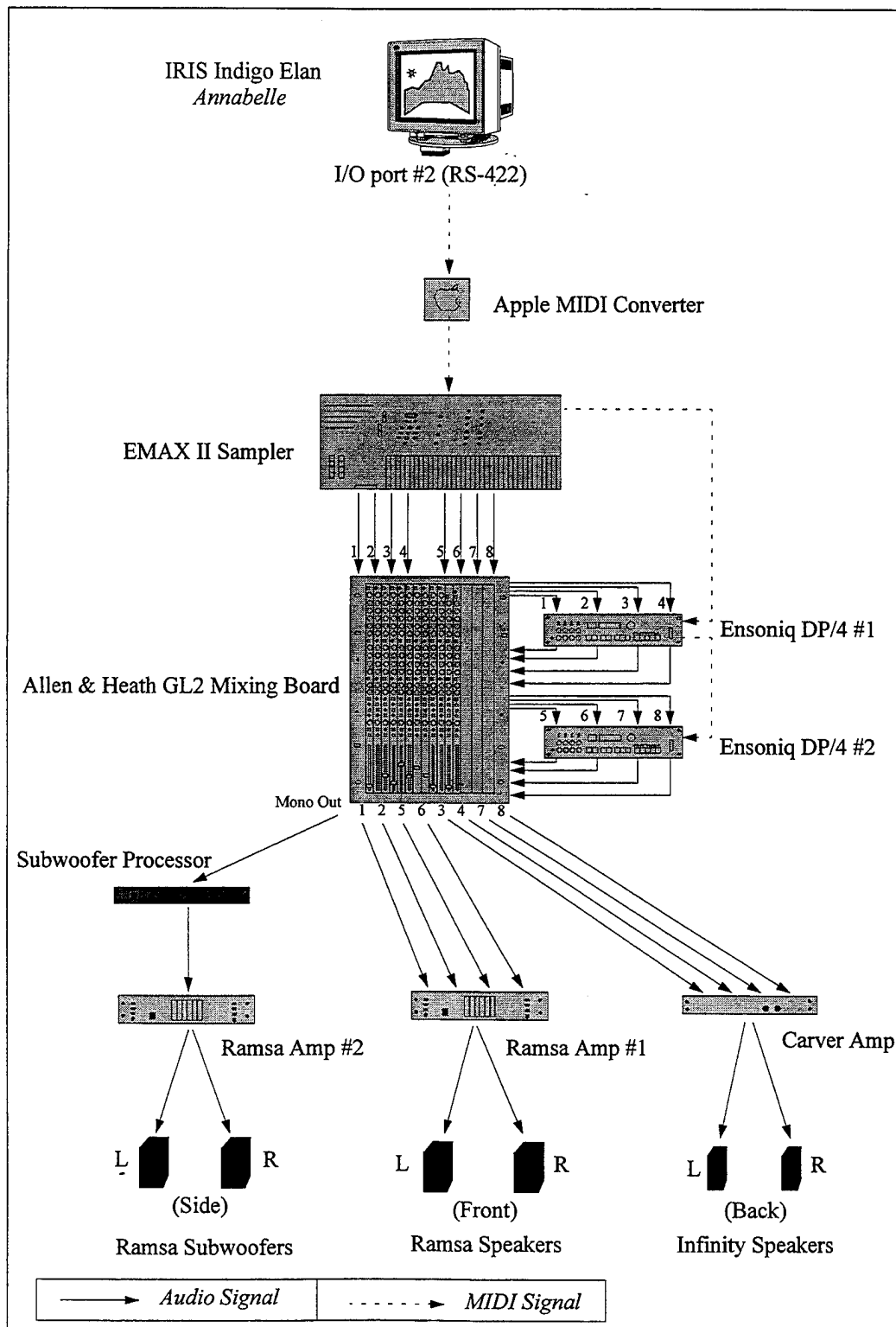
Figure 26: NPSNET-3DSS 2D Hardware Configuration.

Again, the wiring diagrams for this equipment configuration are depicted in APPENDIX C: HARDWARE WIRING DIAGRAMS. The basic hardware configuration to fully implement the SC of the NPSNET-3DSS is depicted in Figure 27.

3. Hardware Flow

The following is a description of the overall hardware flow of NPSNET-3DSS. This hardware flow is identical to both the partial and full SC implementation except as indicated.

a. Computer to Sampler

NPSNET-3DSS uses the same interface as NPSNET-PAS to connect NPSNET with the sound system. The software generates the necessary MIDI commands for output to the second RS-422 communication port (*ttyd2*) on the Iris Indigo Elan. The name of the current Indigo used in this system is *Annabelle*. This signal is then sent to the Apple MIDI Interface which converts the signal from the 8-pin RS-422 format to the 5-pin Deutsche Industri Norm (DIN) MIDI format. This signal is then routed to the *MIDI IN* port on the EMAX II. It should be noted that only MIDI data, not actual sound, is sent to the EMAX II from the Indigo.

b. Sampler to Mixing Board

To run NPSNET-3DSS, the EMAX II sampler must have a specific sound bank loaded into its RAM. This sound bank is loaded by software via a MIDI command during the initialization of running NPSNET-3DSS. This sound bank determines: 1) which sounds can potentially be played, 2) how these sounds are generated, and 3) where the sounds should be generated (i.e. which output ports). This sound bank enables the EMAX II to generate eight independent audio signals which are routed to the Allen & Heath GL2 Mixing Board. A more detailed description on the configuration and use of the EMAX II is contained in APPENDIX D: EMAX II CONFIGURATION AND USE.
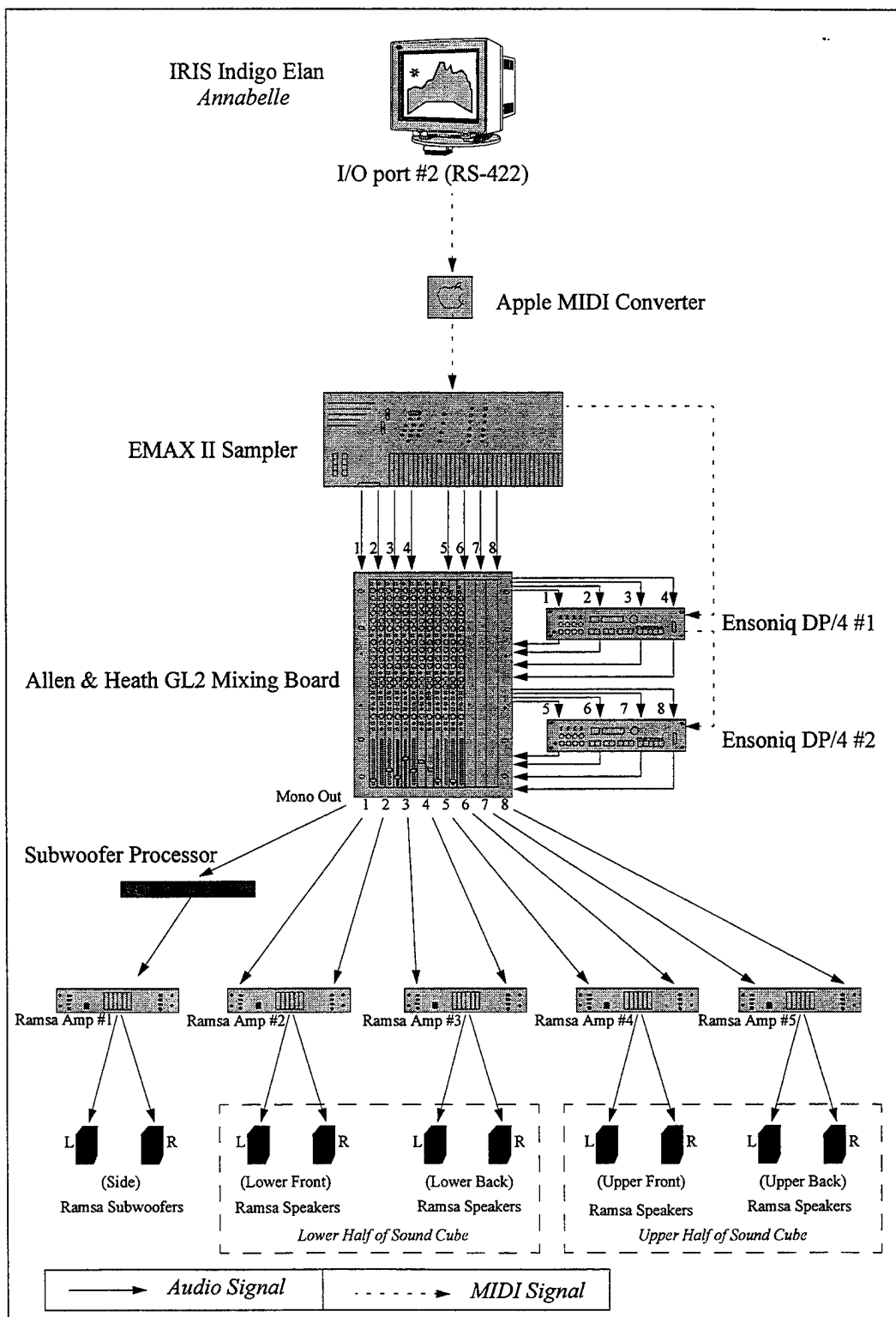
Figure 27: NPSNET-3DSS 3D Hardware Configuration.

c. Mixing Board to and from Digital Signal Processors

The Allen & Heath GL2 Mixing Board is well respected by music engineering aficionados for its extremely clean sound and versatile capabilities. The GL2 receives the eight audio signals from the EMAX II on eight separate audio channels. Each audio channel also has its own insert port to allow routing of the audio signal to and from another audio device. In this case, the other audio device is an Ensoniq DP/4. The DP/4, like the GL2, is a well respected piece of music engineering equipment. Each DP/4 has four independently operating DSPs, and this sound system utilizes two of these DP/4s which provides for a total of eight independently operating DSPs. Each DSP receives one of the eight audio signals sent from the GL2. The audio signal is then processed by the DSP to produce an appropriate amount of reverberation. This processed signal is then returned to the GL2 via the same insert port from which came the original signal. To successfully accomplish this routing of the audio signal from the mixing board to the DSPs, the GL2 and the Ensoniq DP/4s must be configured properly. The process to configure the GL2 is simple (see APPENDIX E: ALLEN & HEATH GL2 MIXING BOARD), but the process to configure the DP/4s is fairly complex and time consuming (see APPENDIX F: ENSONIQ DP/4 DIGITAL SIGNAL PROCESSOR).

d. Mixing Board to Amplifiers/Speakers

The *MONO* output on the GL2 is routed to the Ramsa Subwoofer Processor. The subwoofer processor only boosts the very low frequencies (VLF) of the signal. This VLF is then routed to Ramsa Power Amplifier # 1 for output to both Ramsa Subwoofers. (#1 refers to the current rack mounted position of the Ramsa Amp, where Ramsa Amp #1 is physically located on top of Ramsa Amp #2) Up until now, the hardware flow of both the partial and full SC implementation has been identical, but now there are some differences.

In the partial SC implementation as shown in Figure 26, the audio signals from channels *1, 2, 5*, and *6* on the GL2 are sent to Ramsa Power Amplifier #2 for output to both Ramsa Speakers/Studio Monitors. These audio signals represent the front half of

79

the SC. Accordingly, the audio signals for channels *3, 4, 7,* and *8* on the GL2 are sent to the Carver Power Amplifier for output to the Infinity Reference Speakers. These audio signals represent the back half of the SC. This is how the 3D aspect of the SC is collapsed for use in the current 2D system. As a result, the correct audio signals are being generated for producing the 3D aural cues, but are only amplified in a 2D capable system.

In the full implementation of the SC as shown in Figure 27, the audio signals for channels *1* and *2* on the GL2 are routed to Ramsa Amp #2, and the audio signals for channels *3* and *4* are routed to Ramsa Amp #3. The audio signals *1, 2, 3,* and *4* represent the lower half of the SC. Accordingly, the audio signals for channels *5* and *6* on the GL2 are routed to Ramsa Amp #4, and the audio signals for channels *7* and *8* are routed to Ramsa Amp #5. The audio signals *5, 6, 7,* and *8* represent the upper half of the SC. All the Ramsa Amplifiers in this full SC implementation are routed to a set of Ramsa Speakers. The specifics on how these audio signals are routed from the mixing board to the amplifiers can be found in APPENDIX C: HARDWARE WIRING DIAGRAMS.

XI. IMPLEMENTATION AND ANALYSIS

Thus far, this research effort has been centered primarily around the theory and design of a MIDI-based free-field sound system capable of producing 3D aural cues for use in NPSNET. Given the software and hardware functionality described earlier, this chapter discusses how the 3D Sound Cube Model (SCM), the Precedence Effect (PE) sound model, and synthetic reverberation (SR) are implemented into NPSNET-3DSS. The ultimate goal of this implementation is to increase the effectiveness of the auditory channel in NPSNET by increasing the level of immersion of the NPSNET player.

A.   3D SOUND CUBE MODEL

1. Implementation

Before the 3D SCM could be implemented, the EMAX II had to be completely reconfigured. The previous sound system of NPSNET-PAS used only six of the eight audio outputs on the EMAX II. Thus, the EMAX II was reconfigured with a new sound bank which uses all eight of its audio outputs. This configuration of the EMAX II is explained in greater detail in APPENDIX D: EMAX II CONFIGURATION AND USE. Once the EMAX II was reconfigured having eight independent audio outputs, the signals were routed to the mixing board for eventual output to the speakers. Next, because of the current lack of speakers required for the sound cube (SC) as depicted earlier in Figure 17, the partial SC was temporarily implemented (see Partial Sound Cube Implementation on page 74). Once the partial SC was implemented, the algorithm for the SCM was developed in software using *C++*. The algorithm for the SCM is inherent in the function *trigger_3D_sound* which resides in the file *soundlib.cc*. The code for implementing the SCM algorithm follows directly from the derivation of the SCM described earlier (see 3D SOUND CUBE MODEL on page 57).

2. Analysis

a. Hardware Setup

The 3D SCM was first tested during an NPSNET demonstration using the current 2D sound system, as mentioned earlier, by collapsing the eight speaker positions of the 3D SCM onto the four speakers of the partial SC implementation. For example, in looking back at Figure 21 on page 57, speaker position A was collapsed onto position D, and speaker position B onto C, etc. Thus, any sounds which were to be played in the 3D SCM at positions A and D were simply played independently through one speaker of the 2D sound system at position A.

b. Original SCM

During this first test, the SCM appeared to be working just fine. As sound events occurred in the VE of NPSNET, the NPSNET-3DSS played the proper sound in the proper speakers. This action seemed to indicate that the SCM was in fact producing the proper aural cues allowing the NPSNET player to accurately localize the sound source with the position of the sound event. However, as the test continued, it became apparent that the volume of the sound source was inconsistent at different azimuths relative to the listener while keeping the distance of the sound source constant. For the SCM to work properly, the volume of the sound source should have the same level at the same distance regardless of the azimuth. Something was clearly wrong. As a result, the software and hardware were checked, and finding no problems the test was repeated. Still the problem remained, but during the course of the this second test, the problem was discovered.

The problem occurred when the sound source was located midway between any two sets of speaker positions. In this situation, the SCM evenly distributed the volume of the sound source between these two speakers in the attempt to make the sound appear as though it were emanating at a position midway between the speakers. As a result, the volume of the sound source was reduced by half and played at each of the two speakers. Although the sound did appear to be emanating from a position midway between the two

speakers, the volume was reduced by half. The idea of the SCM was that when both speakers played the sound source at half the volume, the total volume played would then equal that of the original sound source. Although the idea of conserving the total volume of the original sound source looks good in terms of mathematics, this is not how sound works. Thus, the SCM needed to be revised.

c. Revised SCM

In reviewing how the SCM distributes the total volume of the sound source among the speakers, it became clear that the SCM was distributing the wrong volume. The volume which should be distributed is the total volume which potentially can be played through all of the speakers and not just that of the sound source. This research considers the total volume which potentially can be played through all the speakers as the *pool* of volume of the speakers. In other words, if all the speakers were to *pool* the maximum volume that each could generate, the total maximum amount of volume is considered the *pool* of volume. So, if each speaker were to play a sound at its maximum volume level, the resulting apparent location of the sound source would be in the center of the SC.

The basic concept of this revised SCM is identical to that of the original SCM except when distributing the volume to the speakers. In this revised SCM, the volume of the virtual sound source is still calculated using the same psychoacoustically-based law depicted earlier in Eq 14 on page 56. However, the volume of the virtual sound source is now added to each speaker's potential pool of volume. And, it is the total pool of volume which is distributed to the speakers according to the relative location of the virtual sound source with the listener. The total pool of volume of the speakers is a function of the dynamic range of the speakers and room acoustics. Sophisticated speakers having a wide dynamic range will have a larger potential pool of volume. A room with great acoustics will also have a correspondingly greater pool of volume. Thus, the difference between the original SCM and the revised SCM is as follows. In the original SCM, both distance and azimuth aural cues were based on the volume distribution as a result of the relative location

between the virtual sound source and the listener. This approach was not accurate. In the revised SCM, the distance aural cue is only a function of the psychoacoustically-based formula of Eq 14 on page 56, and the azimuth aural cue is a function of distributing the pool of volume based on the relative location between the virtual sound source and the listener. The following is a fragment of code within the function *trigger_3D_sound* in the file *soundlib.cc* which shows how the SCM was revised.

*speaker_volume[q] = (volume + poolvolume/4) + poolvolume * (1 - ((index - 1) angle[q]/sum));*

where,

speaker_volume[] is an array of eight speaker volumes,

q is the index of the speaker volume from 0 to 7 (eight total),

poolvolume is the total pool of volume of the speakers,

index is the number of angles less than 70.5 degrees,

angle[] is an array of angles corresponding to it's speaker location,

sum is the sum of all angles less than 70.5 degrees.

For the sound system used by NPSNET-3DSS, a value of 40 for the pool of volume appears to work very well. However, this value is the result of trial and error during many NPSNET demonstrations, and is still an area of ongoing research.

d. Results

The overall results of using the revised SCM, through numerous NPSNET demonstrations, indicates that the virtual sound source is properly distributed among the speakers of our sound system. As a result, the NPSNET player is given the proper aural cues to localize the virtual sound source with its visual counterpart. Therefore, the 3D SCM produces proper 2D aural localization cues when using the four speakers of the partial SC implementation. Thus, in theory, the 3D SCM is capable of producing 3D aural localization cues when implemented with all eight speakers of the full SC depicted earlier in Figure 17.

B. PRECEDENCE EFFECT SOUND MODEL

1. Implementation

The implementation of the Precedence Effect (PE) sound model uses the same EMAX II configuration and partial SC implementation as described earlier in the 3D SCM implementation. However, the function *trigger_3d_sound* had to be rewritten as well as changing some of the design and functionality of the software which runs NPSNET-3DSS. These changes include modifying the software from having one listening position at the center of the SC, to having eight listening positions correlating to the eight speaker positions of the SC. These eight listening speaker positions are anchored to the NPSNET player's position (the listener). So, when the listener moves around in the VE of NPSNET, the eight listening speaker positions will also move in their offset speaker positions correlating the listener's movement. In looking back at the PE sound model in Figure 22 on page 62, one should see the necessity of keeping track of the location of the speaker positions. When the sound wave intersects a speaker position, we need to generate the sound source at the corresponding speaker. The PE sound model was much simpler to implement than the SCM and it better represents how we hear sounds in the real world.

2. Analysis

Like the SCM, the PE sound model was tested via NPSNET. Although the PE sound model is easily implemented and more accurately reflects our perception of sound, it was not effective, for it could not generate all eight sounds to the speakers within 30 milliseconds (see The Precedence Effect on page 24). The reason for its ineffectiveness lies in the delay of communication signals associated with MIDI. This is commonly referred to as MIDI delay, and has been a constant source of trouble for music engineers in their attempt to synchronize numerous tracks on a sequence. This MIDI delay, which is indeed a real communication problem, is part of the MIDI Specification. Specifically, the MIDI Specification says the following: "The [MIDI] interface operates at 31.25 (+/- 1%) kbaud,

asynchronous, with a start bit, 8 data bits (D0 to D7), and a stop bit. This makes a total of 10 bits for a period of 320 microseconds per serial byte" [INTE83]. At first glance, 320 microseconds seems well under the 30 milliseconds constraint of the PE. However, MIDI commands are sent in blocks of three specific commands. And, to play a discrete sound in NPSNET-3DSS, such as an explosion, we need to send three MIDI commands to play the note associated with the explosion. Next, we need to send three MIDI commands to stop playing the same note. In essence, we turn on the note and then we turn off the note. The following is a fragment of code in the file *soundlib.cc* which shows how these MIDI commands are sent.

```
send_midi_command( midiport, (unsigned char) (NOTE_ON + channel));
send_midi_command( midiport, (unsigned char) sound);
send_midi_command( midiport, (unsigned char) volume);

send_midi_command( midiport, (unsigned char) (NOTE_OFF + channel));
send_midi_command( midiport, (unsigned char) sound);
send_midi_command( midiport, (unsigned char) 0);
```

The *midiport* identifies which RS-422 port on the Iris workstation to send the MIDI commands and is not a key factor in the MIDI delay. The remaining commands prefaced by the type conversion *(unsigned char)* such as *(NOTE_ON + channel)*, *sound*, and *volume* are the MIDI commands sent out to turn on/off the particular note. Each of these commands consists of two bytes which makes a total of twelve bytes to turn on and turn off a note. Since each byte takes 320 microseconds to send, it then takes 3.84 milliseconds to send these twelve bytes to turn on and off one sound. But we have eight independent sounds to generate, so it will take 46 milliseconds to generate all eight sounds. This exceeds the 30 millisecond constraint of the PE, hence rendering the PE sound model ineffective. As a result, when running NPSNET-3DSS with the PE sound model, it was impossible to localize any sound sources. For when any single sound event occurred, the perception heard by the listener what that of multiple sounds emanating from multiple directions rendering localization impossible.

C.   SYNTHETIC REVERBERATION

1. Implementation

a. Hardware Setup

The same EMAX II configuration and partial SC implementation as described earlier are used when implementing synthetic reverberation (SR). To generate the SR, a digital signal processor (DSP) is needed (see Application in Virtual Environments on page 68). The DSP used in this research is the Ensoniq DP/4 Parallel Effects Processor [ENSO92a] [ENSO92b]. Each DP/4 has four independent processors labeled A, B, C, and D which can be programmed individually. The basic idea in using the DP/4s is to allocate one processor for each audio channel which is in turn routed to each speaker. As a result, NPSNET-3DSS utilizes two DP/4s. Looking back at Figure 26 on page 76, we can see how the DP/4s interface with the sound system for use in partial SC implementation. But before the DP/4s can be used to generate any form of SR, they need to be preprogrammed in an appropriate configuration. For a detailed description on how to configure the DP/4s for use in NPSNET-3DSS, see APPENDIX F: ENSONIQ DP/4 DIGITAL SIGNAL PROCESSOR. Furthermore, to access the DP/4s via MIDI, the function *trigger_3d_sound* was modified to add the SR functionality.

b. Ambient Environment

Once the DP/4s have been preprogrammed with the desired reverberation algorithm, they can now be used to generate the SR as a function of the early echoes caused by reflections. The number and amplitude of these reflections is based on the listener's position in the virtual world. As mentioned earlier, a bounding volume encasing a specific area having a certain desired reverberation effect (i.e. a valley, or canyon, etc.), can be created from the x, y, and z coordinates within the virtual world. So, when the listener enters this bounding volume, a MIDI command is sent to the DP/4s instructing them to change to

87

a new reverberation algorithm. This procedure was actually the last feature implemented in NPSNET-PAS [ROES94]. To change reverberation algorithms, this procedure was based on sending MIDI program change information to the DP/4s which in turn loaded a new reverberation algorithm into the processors of the DP/4. Although effective in changing reverberation algorithms, it was done in real-time. This is one of the few faults of the DP/4, for when the DP/4 reloads any type of new algorithm into any one or all of its processors, all sounds routed through the DP/4 stop until the new algorithm is loaded. The amount of delay varies based on the particular algorithm selected and in how many processors the algorithm will be reloaded. In talking to a representative of the Ensoniq Corporation, the makers of the DP/4, it was discovered that they were aware of the problem and it was corrected in their updated product the Ensoniq DP/4 Plus.

To correct the delay problem when switching algorithms, a new method for switching reverberation algorithms in real-time was implemented. The solution to this problem is that we do not switch the algorithms. Instead we keep the same reverberation algorithm loaded, and we simply change certain parameters of the algorithm via real-time MIDI modulation messages. These real-time modulation messages follow a specific format as described in the MIDI Specification (see [INTE83]). The basic idea is to map a specific MIDI modulation message to the particular parameter of the reverberation algorithm that is to be changed in real-time based on the listener's position in the virtual world. Accordingly, the DP/4 must be preprogrammed to recognize which of these MIDI modulation messages will control the specific parameters of the already loaded reverberation algorithm (see [ENSO92a] [ENSO92b]). After trial and error and after consulting with the Ensoniq Corporation, the *Large Room Rev* algorithm was selected as the best overall algorithm to use in NPSNET for it provides a wide range of reverberation and decay. Thus, part of the initialization process when NPSNET-3DSS is started is to load the *Large Room Rev* algorithm into all four processors of both DP/4s. To understand how these algorithms are loaded, refer to APPENDIX F: ENSONIQ DP/4 DIGITAL SIGNAL

PROCESSOR. However, the basic idea is to allocate a MIDI channel for each processor and then send the MIDI command on the processor's MIDI channel which loads the appropriate algorithm. The following is the portion of code within the file *soundlib.cc* which loads these algorithms.

```
//load top DP/4
//load algorithm in processor A
send_midi_command( midiPort, (unsigned char) 0xC0);
send_midi_command( midiPort, (unsigned char) 0x00);

//load algorithm in processor B
send_midi_command( midiPort, (unsigned char) 0xC1);
send_midi_command( midiPort, (unsigned char) 0x01);

//load algorithm in processor C
send_midi_command( midiPort, (unsigned char) 0xC2);
send_midi_command( midiPort, (unsigned char) 0x02);

//load algorithm in processor D
send_midi_command( midiPort, (unsigned char) 0xC3);
send_midi_command( midiPort, (unsigned char) 0x03);

//load bottom DP/4
//load algorithm in processor A
send_midi_command( midiPort, (unsigned char) 0xC6);
send_midi_command( midiPort, (unsigned char) 0x00);

//load algorithm in processor B
send_midi_command( midiPort, (unsigned char) 0xC7);
send_midi_command( midiPort, (unsigned char) 0x01);

//load algorithm in processor C
send_midi_command( midiPort, (unsigned char) 0xC8);
send_midi_command( midiPort, (unsigned char) 0x02);

//load algorithm in processor D
send_midi_command( midiPort, (unsigned char) 0xC9);
send_midi_command( midiPort, (unsigned char) 0x03);
```

The *Large Room Rev* algorithm has twenty-two parameters that potentially can be changed in real-time to produce virtually any type of reverberation effect desired. But, as with all the algorithms of the DP/4, only two of these parameters can be assigned MIDI modulation messages. Thus, it is important to consider which two parameters to utilize, for all future potential reverberation effects will be based on these two chosen

parameters. Again, after trail and error and consulting with Ensoniq Corporation, the two parameters chosen were *03 Room/Hall Decay* and *06 Room/Hall HF Damping*. Like the decision to use the *Large Room Rev* algorithm, these parameters were chosen for they offer adequate reverberation cues for use in NPSNET. However, this research effort is focused on the feasibility and practicality of using commercial off-the-shelf equipment like the DP/4 for use in VE applications, and is not a research in the analysis/development of producing SR algorithms. As a result, more research needs to be done in identifying the optimal algorithms (factory presets or customized) and possible parameters to utilize for generating the greatest perceptual effects when using SR in the VE of NPSNET.

Now that we have identified how to generate SR to recreate various ambient conditions, all that remains is creating the bounding volumes encompassing the desired SR effect based on the listener's position in the virtual world.

c. *Distance Perception*

Because of the real-time constraint, we cannot switch algorithms in the DP/4s between the use of SR for recreating ambient conditions and the use of generating an increased perception of distance. Thus, the choice of using the *Large Room Rev* algorithm and the *03 Room/Hall Decay* and *06 Room/Hall HF Damping* parameters was not only dependent on the use of SR for recreating ambient conditions, but also on the use of SR for generating an increased perception of distance. So, the same algorithm and parameters that are used to recreate ambient conditions are used to generate the SR needed for increased distance perception.

Implementing the distance perception cues is simply a function of the distance between the listener and the sound source. As the distance increases, so does the decay of the sound source increase as a result of the echoes caused by reflections. Likewise, as the distance increases, so does the HF damping of the sound source increase. Therefore, what is needed, is a mapping of the distance to the amount of decay and HF damping required to produce the appropriate SR for generating the perception of increased distance.

As mentioned earlier, the focus of this research is on the feasibility of using off-the-shelf equipment for use in VE applications, and not the actual implementation of acoustically accurate SR algorithms. As such, a basic algorithm was developed by playing a sound source at a known distance, and then applying a certain amount of SR to the audio signal. This procedure was repeated at numerous distances from zero to eight hundred meters using various amounts of decay and HF damping. Eventually, an algorithm evolved which sends appropriate MIDI modulation messages to the DP/4s for generating the required SR needed for increased distance perception. The following is part of the code which can be found in the file *soundlib.cc* which produces the distance perception SR in one of the DP/ 4s at a distance between 50 and 99 meters.

```
//change amount of decay in processor A
send_midi_command( midiPort, (unsigned char) 0xB5);
send_midi_command( midiPort, (unsigned char) 0x0B);
send_midi_command( midiPort, (unsigned char) 0x10);

//change amount of HF damping in processor A
send_midi_command( midiPort, (unsigned char) 0xB5);
send_midi_command( midiPort, (unsigned char) 0x0C);
send_midi_command( midiPort, (unsigned char) 0x10);

//change amount of decay in processor B
send_midi_command( midiPort, (unsigned char) 0xB5);
send_midi_command( midiPort, (unsigned char) 0x0D);
send_midi_command( midiPort, (unsigned char) 0x10);

//change amount of HF damping in processor B
send_midi_command( midiPort, (unsigned char) 0xB5);
send_midi_command( midiPort, (unsigned char) 0x0E);
send_midi_command( midiPort, (unsigned char) 0x10);

//change amount of decay in processor C
send_midi_command( midiPort, (unsigned char) 0xB5);
send_midi_command( midiPort, (unsigned char) 0x0F);
send_midi_command( midiPort, (unsigned char) 0x10);

//change amount of HF damping in processor C
send_midi_command( midiPort, (unsigned char) 0xB5);
send_midi_command( midiPort, (unsigned char) 0x10);
send_midi_command( midiPort, (unsigned char) 0x10);

//change amount of decay in processor D
send_midi_command( midiPort, (unsigned char) 0xB5);
```

```
send_midi_command( midiPort, (unsigned char) 0x11);
send_midi_command( midiPort, (unsigned char) 0x10);

//change amount of HF damping in processor D
send_midi_command( midiPort, (unsigned char) 0xB5);
send_midi_command( midiPort, (unsigned char) 0x12);
send_midi_command( midiPort, (unsigned char) 0x10);
```

2. Analysis

Once the ability to use SR in real-time was implemented, this research effort focused on fine tuning the use of SR to enhance the listener's distance perception of sound events as opposed to recreating ambient conditions in NPSNET. The following describes why distance perception was emphasized as opposed to recreating ambient conditions.

a. Ambient Environment

The reason for not focusing on recreating ambient conditions is that this research effort is oriented towards immersion into NPSNET through some sort of vehicle (i.e. tank or helicopter). Currently, in typical NPSNET scenarios, tanks and helicopters operate in fairly consistent acoustic environments. Thus, there are not too many opportunities to provide the listener with different ambient cues. Granted there are times where the listener's ambient environment will change while sitting inside a vehicle, but for the most part the ambient conditions to the listener will be fairly consistent. For example, when a helicopter is flying around, it is usually not flying through many types of different acoustic conditions. Conversely, during virtually all NPSNET scenarios, there are numerous weapons being fired and explosions impacting all around the listener. As a result, there are many opportunities for which SR can be applied to help increase the perception of distance of these numerous sound events. Therefore, to gain the most from the auditory channel in the current scenarios of NPSNET, the use of SR for increased distance perception is emphasized over recreating ambient conditions.

Another reason for not focusing on recreating ambient conditions is that the goal of developing the procedure to recreate ambient conditions in real-time has been realized. All that remains now is to define more bounding volumes like those already

proven effective in NPSNET-PAS. The result will be an acoustic mapping of NPSNET based on the bounding volumes of x, y, and z world coordinates for encompassing the desired SR effects. Like in NPSNET-PAS, when a listener is inside these bounding volumes, the particular MIDI modulation messages can be sent to the DP/4s to generate the necessary SR for creating the desired ambient environment.

b. Distance Perception

The effectiveness of the DP/4s for generating the required SR needed for increased distance perception was tested during typical NPSNET scenarios. The results showed that the DP/4s could adequately provide SR in real-time by using MIDI modulation messages. As the distance from the listener to the sound source increases, there is a noticeable increase in the decay time of the sound source and the sound source is more muffled as a result of the increased HF damping. Furthermore, when the use of SR is coupled with the visual cue, the distance perception of the sound source becomes more pronounced than in the previous NPSNET sound systems. In these previous systems, the only aural cue for judging distance was volume. However, in this sound system the listener is not only provided the aural cue of volume, but also the aural cues of reverberation and decay to help judge distance. Further analysis, though, indicates that the *Large Room Rev* algorithm needs to be modified or replaced so that the SR produced is more similar to that of outdoors reverberation. However, finding an appropriate outdoor reverberation algorithm may or may not be possible because of all the uncontrolled permutations associated with outdoor acoustics. Another factor to be considered is that of the sampled sounds themselves. Perhaps better quality sampled sounds are needed to reproduce better quality SR. Additionally, the current algorithm which determines how much SR to produce is based on discrete distances of fifty meters out to eight hundred meters. This algorithm needs to be changed so that the SR produced is determined via an analog algorithm based on any amount of distance from the listener to the maximum range of the sound source and not discrete fifty meter intervals. Nevertheless, the goal of using the DP/4s to generate the

required SR for increased distance perception is realized, thus providing for a more realistic acoustic environment for the NPSNET player.

XII. CONCLUSION

A.   OVERALL RESULTS

The overall result of this research effort is a MIDI-based free-field sound system, NPSNET-3DSS, consisting of off-the-shelf sound equipment and computer software capable of generating aural cues in three dimensions for use in the VE of NPSNET. NPSNET-3DSS has been tested during numerous demonstrations of NPSNET and has proved capable of generating SR for increased distance perception and the eight independent audio channels required for potential output to a cube-like configuration of eight loudspeakers. This research effort lays the foundation for increasing one's level of immersion in NPSNET through effective use of the auditory channel.

B.   FOLLOW-ON WORK

Although this research effort has improved the effectiveness of the auditory channel for use in NPSNET, there remains much work to be done. The following are some possible areas of follow-on work.

1. Sound Cube

It is important to note that the speakers identified for use in the full SC implementation are all of the same type -- Ramsa WS-A200. The reason for having the same type speakers in to ensure that all speakers are matched properly in phase with each other. If the speakers are not properly matched, the spatial effect of the 3D cues will be severely degraded. Hence, the importance of using properly matched speakers cannot be undermined. Also, the use of *The Ultimate Speaker Stand* is recommended to support the upper four speakers of the SC.

Since NPSNET-3DSS is already generating audio as if there were a full SC, all that is needed is the additional amplifiers and speakers to fully implement the SC. Upon arrival of this equipment, one simply has to route the appropriate audio cables to this equipment

and orient the speakers in the SC configuration. When the SC is implemented, it will not only provide 3D aural cues for use in NPSNET, but will also function as a valuable research tool for further investigations on the use of free-field systems for virtually any audio application.

2. Ambient Sounds

Ambient sounds produce an enormous amount of aural cues which in turn helps the listener to identify the surrounding environment. Some of these ambient sounds are very indicative of particular environments. For example, the sounds of the city are vastly different than that of the jungle. Also, when we are in the city or the jungle, we rarely single out a certain sound in an attempt to localize the sound, unless of course, a police car with it's siren sounding is whizzing by within our visual acuity. For the most part, because there are so many sounds and of so many varieties, we normally listen to these sounds as a group -- the ambient sound. Thus, adding the appropriate ambient sounds to a VE will no doubt greatly increase one's immersion within that VE, as opposed to having no ambient sounds. The idea is to capture the ambient sounds typical of our VE. This can be done by using a DAT recorder and actually recording the sounds while physically located in the environment whose ambient sounds we want to capture. Or, we can purchase prerecorded ambient sounds for virtually any type of environment from any one of numerous commercial vendors. Both of these options are now available for use in future research. A JVC portable DAT recorder has recently been purchased for use by the NRG which can be used to record not only specifically intended sounds but also ambient sounds. And, a collection of numerous ambient sounds produced by *Sound Ideas* has also recently been purchased for use by the NRG.

A piece of sound equipment recently purchased by the NRG is the *Lexicon CP-1 Plus Digital Audio Environment Processor* (see APPENDIX H: SOUND PERCEPTION EXPERIMENTS). Lexicon is well respected in the musical world for having some of the best reverberation algorithms. The CP-1 Plus has the capability of recreating various types

of ambient conditions. As a result, prerecorded sounds can be sent to the CP-1 Plus and then processed to produce the desired ambient effect.

A feature of the CP-1 Plus that has great potential is that of the binaural recording mode. This mode processes binaural recording signals, which are intended for headphone listening (see APPENDIX G: BINAURAL RECORDINGS), and presents them via loudspeakers. As a result of having done some preliminary experiments with the CP-1 Plus using binaural recordings of ambient sounds, the effect produced by the resulting processed ambient sound is remarkable. The dynamic range of the processed sound was quite large recreating a very convincing ambient environment. Because of the binaural mode, the CP-1 Plus acts as a bridge between headphone and free-field systems. There is indeed great research potential with the CP-1 Plus.

3. Headphone System

All previous NPSNET sound servers have focused on the generation of aural cues via free-field format. The technological state of digital signal processing and microprocessors was probably the primary reason for the bias towards using free-field systems to date. However, today's DSPs and CPUs are extremely powerful offering capabilities well ahead of their predecessors. The computational power required for headphone systems can now tap the power of these DSPs and CPUs. Thus, the time has come for the development of a headphone delivery system for use in NPSNET.

4. Hybrid Sound Delivery System

In a group meeting at NASA-Ames Research Center with Durand Begault, Elizabeth Wenzel, Brent Gillespie (from CCRMA), I started a discussion on the advantages and disadvantages of headphone and free-field delivery systems. One of the interesting points brought out in this discussion was that of a hybrid sound system for use with VEs consisting of both headphones and loudspeakers. The headphones can be used in conjunction with a motion tracker such as a Polhemus Fastrack to generate certain aural cues to the listener critical to head motion. The loudspeakers can focus on generating

ambient sounds as well as the VLF that the headphones are incapable of generating. The result is a sound system that maximizes the advantages and minimizes the disadvantages of each sound system. Whatever the exact role of each sound system, the potential effectiveness of this hybrid sound system warrants further research.

C. RECOMMENDATIONS

There are numerous recommendations which can be made to help improve the development of future sound systems for NPSNET. The following are few of the more pertinent recommendations.

1. Audio Research Environment

As discussed earlier in Chapter 1, the current working environment used to do research and development of audio applications for use in NPSNET lacks access to an anechoic chamber, common electrical ground, and continuity of audio expertise. To increase the potential success of future research and development, these limiting factors must be eliminated. Furthermore, a library of sound related references should also be made available within this working environment for ease of use and immediate access for future research and development. Although the current area utilized for developing NPSNET-3DSS has been improved over the course of this research effort, improvements are still needed and upgrades of sound related hardware and software must always be considered.

2. Simplified Sound System

Even though NPSNET-3DSS adequately provides aural cues for use in NPSNET, it is nevertheless comprised of numerous types of sound related hardware and software. In order to make future sound systems more portable and standardized, it is recommended to consider moving the bulk of this sound hardware and software to a more simplified system perhaps comprising a single vendor for a kind of *one-stop-shop* sound system. A possible choice of venders is SGI, not only because all the graphic workstations used in the

development of NPSNET are SGI machines, but also because of the recent advances and future developments of SGI audio applications.

3. New Computer Audio Course

Because of the recent advances in computer audio applications, more people are becoming exposed to computer audio resulting in more users of computer audio. However, there is no course in any type of computer audio offered in the computer science curriculum at NPS. Students just manage to find a way to apply audio in their projects without any instruction as to how computer audio works and the correct ways to use computer audio. It is recommended that some sort of computer audio instruction be offered at NPS as a stand alone course or perhaps as part of a multimedia course.

4. Multi-Modal Thinking

To increase the level of immersion of future NPSNET applications, we must start thinking in terms of the multi-modal aspects of NPSNET. For example, the primary focus of the NRG has been on the enhancement of the visual channel of the VE of NPSNET, and just recently efforts have been made to enhance the effectiveness of the audio channel. Soon, no doubt, enhancements will need to be made in the area of haptics (perhaps this should be called the *haptic channel*). The point being, we cannot continue to look at each mode (visual, audio, and haptic) as a separate aspect to be enhanced. We must start considering how each mode effects the other when integrated together for the purpose of increasing one's immersion into virtual worlds.

5. The Artistic Aspect of Sound

Although much work has been done integrating sound for use in NPSNET, the focus has been purely scientific. As such, the work done thus far in applying aural cues for use in NPSNET is devoid of any artistic qualities. In order to broaden and perhaps improve the quality of audio applications in future NPSNET sound systems, we must start considering the artistic aspects of sound.

D. FINAL THOUGHTS

Probably the most important aspect of this research effort has been to not only provide insights into the past design decisions of previous NPSNET sound systems, but also to provide direction for future NPSNET sound systems. It is hoped that this research effort will not only help to establish the NRG as a leader in the application of 3D sound for use in VEs, but will also help to establish of the necessity for a permanent computer audio research facility within the Department of Computer Science at NPS.

LIST OF REFERENCES

[BALL91]   Ballou, G., *Handbook for Sound Engineers*, 2nd Ed., Howard W. Sams & Company, Carmel, Indiana, 1991.

[BEGA91]   Begault, D. R., "Preferred Sound Intensity Increase for Sensation of Half Distance," *Perceptual and Motor Skills*, Vol. 72, pp. 1019-1029, 1991.

[BEGA92]   Begault, D. R., "Perceptual Effects of Synthetic Reverberation on Three-Dimensional Audio Systems," *Journal of the Audio Engineering Society*, Vol. 40, No. 11, pp. 895-904, November, 1992.

[BEGA94]   Begault, D. R., *3-D Sound for Virtual Reality and Multimedia*, Academic Press, Cambridge, Massachusetts, 1994.

[BLAU83]   Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, Massachusetts, 1983.

[BREG90]   Bregman, A. S., *Auditory Scene Analysis*, MIT Press, Cambridge, MA, 1990.

[BRUN93]   Brungart, D. S., "Distance Simulation in Virtual Acoustic Displays," in *Proceedings of the IEEE 1993 National Aerospace and Electronics Conference. NAECON 1993*, Vol. 2, pp. 612-617, 1993.

[BURG92]   Burgess, D., *Real-Time Audio Spatialization with Inexpensive Hardware*, Georgia Institute of Technology, October, 1992.

[CHOW82]   Chowning, J. and Sheeline, C., *Auditory Distance Perception Under Natural Sounding Conditions*, Report No. STAN-M-12, Department of Music, Center for Computer Research in Musics and Acoustics (CCRMA), Stanford University, November, 1982.

[DAHL92]   Dahl, L., *NPSNET: Aural Cues For Virtual World Immersion*, Master of Computer Science Thesis, Naval Postgraduate School, September, 1992.

[DEER89]   Deering, S., "Host Extensions for IP Multicasting," *RFC 1112*, August 1989.

[DIGI90]   Digidesign Incorporated, *Sound Designer II*, Version 2.0, Digidesign Incorporated, 1990.

[DOAN94]   Doan, T., "Understanding MIDI," *IEEE Potentials*, Vol. 13, pp. 10-11, February, 1994.

[DUDA95] Duda, R., *3-D Sound Perception*, Presented during the CCRMA Summer Workshop: Introduction to Psychoacoustics and Psychophysics with emphasis on the audio and haptic components of virtual reality design, Stanford University, June 26 - July 8, 1995.

[DURL95] Durlach, N., Mavor, A., (Eds.), *Virtual Reality, Scientific and Technological Challenges*, National Academy Press, Washington, D.C., 1995.

[EMU89] E-mu Systems, Incorporated, *EMAX II 16-bit Digital Sound System Operation Manual*, E-mu Systems Inc., 1989.

[ENSO92a] Ensoniq Corporation, *DP/4 Musician's Manual*, Version 1.02, Ensoniq Corporation, 1992.

[ENSO92b] Ensoniq Corporation, *DP/4 MIDI SysEx Implementation Specification*, Version 1.0, Ensoniq Corporation, 1992.

[ERIC93] Ericson, M., D'Angelo, W., Scarborough, E., Rodgers, S., Amburn, P., and Ruck, D., "Applications of Virtual Audio," *Proceedings of the IEEE 1993 National Aerospace and Electronics Conference. NAECON 1993*, Vol 2., pp. 604-611, 1993.

[EVAN93] Evans, B., "Enhancing Scientific Animations with Sonic Maps: An Introduction to Data Sonification," Course 81 Notes, pp. 3.5-3.12, ACM SIGGRAPH '93.

[EVER91a] Everest, F., "Fundamentals of Sound," in G. Ballou (Ed.) *Handbook for Sound Engineers*, 2nd Ed., pp. 3-24, Howard W. Sams & Company, Carmel, Indiana, 1991.

[EVER91b] Everest, F., "Psychoacoustics," in G. Ballou (Ed.) *Handbook for Sound Engineers*, 2nd Ed., pp. 25-42, Howard W. Sams & Company, Carmel, Indiana, 1991.

[FARA90] Farallon Computing Incorporated, *MacRecorder Sound System*, Farallon Computing Incorporated, 1990.

[GARD73] Gardner, M., Gardner, R., "Problem of Localization in the Median Plane: Effect of Pinnae Cavity Occlusion," *Journal of the Acoustical Society of America*, Vol. 53, pp. 400-408, 1973.

[GILL95a] Gillespie, B., *Convolution and the HRTF*, Presented during the CCRMA Summer Workshop: Introduction to Psychoacoustics and Psychophysics with emphasis on the audio and haptic components of virtual reality design, Stanford University, June 26 - July 8, 1995.

[GILL95b]   Gillespie, B., *Wave Physics*, Presented during the CCRMA Summer Workshop: Introduction to Psychoacoustics and Psychophysics with emphasis on the audio and haptic components of virtual reality design, Stanford University, June 26 - July 8, 1995.

[GILL95c]   Gillespie, B., *Localization and Echo Experiments*, Presented during the CCRMA Summer Workshop: Introduction to Psychoacoustics and Psychophysics with emphasis on the audio and haptic components of virtual reality design, Stanford University, June 26 - July 8, 1995.

[HENR91]   Henricksen, C., "Loudspeakers, Enclosures, and Headphones," in G. Ballou (Ed.) *Handbook for Sound Engineers*, 2nd Ed., pp. 497-591, Howard W. Sams & Company, Carmel, Indiana, 1991.

[IEEE93]   Institute of Electrical and Electronics Engineers, International Standard, ANSI/IEEE Std. 1278-1993, *Standard for Information Technology, Protocols for Distributed Interactive Simulation*, March, 1993.

[INTE83]   International MIDI Association, *1.0 MIDI Specification*, copyright 1983.

[INST93]   Institute for Simulation and Training, *Standard for Information Technology - Protocols for Distributed Interactive Simulation Applications*, Version 2.0 Third Draft, IST-PD-90-2, Orlando, Florida, May, 1993.

[LEHR91]   Lehrman, P., and Tully, T., "Catch a Wave: Digital Audio," *MacUser*, October, 1991.

[LORD07]   Lord Rayleigh, Strutt, J., "On Our Perception of Sound Direction," *Philosophical Magazine*, Vol. 13, pp. 214-232, 1907.

[MACE94]   Macedonia, M., Zyda, M., Pratt, D., Barham, P. and Zeswitz, S., "NPSNET: A Network Software Architecture for Large Scale Virtual Environments," *Presence*, Vol. 3, No. 4, pp. 265-287, Fall, 1994.

[MART92]   Martens, W., *Demystifying Spatial Audio*, Ono-Sendai Corporation, 1992.

[MCMI94]   McMillen, K., Wessel, D. and Wright, M., "The ZIPI Music Parameter Description Language," *Computer Music Journal*, Vol 18, Winter, 1994.

[MILL72]   Mills, A., *Auditory Localization, Foundations of Modern Auditory Theory*, Vol. II, Academic Press, New York, New York, 1972.

[MOOR79]   Moorer, J. A., "About This Reverberation Business," *Computer Music Journal*, Vol. 3, Number 2, pp. 13-28, 1979.

[ODON91]   O'Donnell, B., "What is MIDI, Anyway?," *Electronic Musician*, pp.74, January, 1991.

[OLDF84]   Oldfield, S. and Parker, S., "Acuity of Sound Localization: A Topography of Auditory Space. II. Pinna Cues Absent," *Perception*, Vol. 13, pp. 601-617, 1984.

[OPCO90]   Opcode Systems Incorporated, Studio Vision Integrated MIDI and Digital Audio Recording, Version 1.3, Opcode Systems Incorporated, 1990.

[PLEN74]   Plenge, G., "On the Differences Between Localization and Lateralization," *Journal of the Acoustical Society of America.*, Vol. 56, pp. 944-951, 1974.

[ROES94]   Roesli, J., *Free-Field Spatialized Aural Cues for Synthetic Environments*, Master of Computer Science Thesis, Naval Postgraduate School, September, 1994.

[SABI72]   Sabine, W. C., "Reverberation," in R. B. Lindsay (Ed.) *Acoustics: Historical and Philosophical Development*, Dowden, Hutchinson, and Ross, Stroudsburg, Pennsylvania, 1972.

[SAPP95]   Sapp, C., *The Decibel*, Presented during the CCRMA Summer Workshop: Introduction to Psychoacoustics and Psychophysics with emphasis on the audio and haptic components of virtual reality design, Stanford University, June 26 - July 8, 1995.

[SCHR61]   Schroeder, M. R., "Improved Quasi-Stereophony and Colorless Artificial Reverberation," *Journal of the Audio Engineering Society*, Vol. 33, p. 1061, 1961.

[SCHR62]   Schroeder, M. R., "Natural-Sounding Artificial Reverberation," *Journal of the Audio Engineering Society*, Vol. 10, pp. 219-223, 1962.

[SHAW74]   Shaw, E., "The External Ear," *Handbook of Sensory Physiology*, Vol. V/1, Auditory System, (Eds.) Keidel, W. and Neff, W., pp. 455-490, Springer-Verlag, New York, 1974.

[TAGH94]   Taghavy, D., "Take a Roller Coaster Ride in Rolands's Sound Space," *Electronic Musician*, August,1994.

[TONN94]   Tonnesen, C. and Steinmetz, J., "3D Sound Synthesis," *Encyclopedia of Virtual Environments Homepage* (http://gimble.cs.umd.edu/vrtp/eve-main.html), Department of Computer Science, University of Maryland, 1994.

[WENZ90]   Wenzel, E. M., Begault, D. R., *Techniques and Applications for Binaural Sound Manipulation in Human-Machine Interfaces*, NASA-Ames Research Center, Moffett Field, California, 1990.

[WENZ92]   Wenzel, E. M., "Localization in Virtual Acoustic Displays," *Presence*, Vol. 1, No. 1, pp.80, Winter, 1992.

[WENZ95]   Wenzel, E.M., *Sound Localization*, Presented during the CCRMA Summer Workshop: Introduction to Psychoacoustics and Psychophysics with emphasis on the audio and haptic components of virtual reality design, Stanford University, June 26 - July 8, 1995.

[WILK93]   Wilkinson, Scott; "The Thrill of Adventure," *Electronic Musician*, pp. 22, June, 1993.

[WILL76]   Williams, J, Trinklein, F. and Metcalfe, H., *Modern Physics*, Holt, Rinehart and Winston Publishers, New York, 1976.

[ZYDA93]   Zyda, M., Pratt, D., Falby, J., Barham, P. and Kelleher, K., "NPSNET and the Naval Postgraduate School Graphics and Video Laboratory," *Presence*, Vol. 2, No. 3, pp. 244-258, Summer, 1993.

[ZYDA94]   Zyda, M., Pratt, D., Falby, J., Lombardo, C. and Kelleher, K., "The Software Required for the Computer Generation of Virtual Environments," *Presence*, Vol. 2, No. 2, pp. 130-140, Spring, 1993.

BIBLIOGRAPHY

Allen, J. B., and Berkeley, D. A., "Image Model for Efficiently Modeling Small-Room Acoustics", *Journal of the Acoustical Society of America*, Vol. 65, pp. 943-950, 1979.

Ando, Y., *Concert Hall Acoustics,* Berlin: Springer-Verlag, 1985.

Asano, F., Suzuki, Y., and Stone, T., "Role of spectral cues in median plane localization", *Journal of the Acoustical Society of America*, Vol. 88, pp. 159-168, 1990.

Backus, J., *The Acoustical Foundations of Music,* W. W. Norton & Company, New York, 1977. (Wave physics fundamentals of music.)

Ballou, G. (Ed.), *Handbook for Sound Engineers: The New Audio Cyclopedia,* Howard W. Sames & Co., Carmel, Indiana, 1991. (A great source of much sound related information.)

Bauck, J. and Cooper, D., "Generalized Transaural Stereo", in *Proceedings of the 93rd Convention of the Audio Engineering Society*, San Francisco, CA, October, 1992. (Shows how to get 3D audio from two loudspeakers. See also Klayman)

Begault, D. R. & Wenzel, E. M., "Techniques and applications for binaural sound manipulation in man-machine interfaces," *NASA TM102279*, 1990.

Begault, D. R. & Wenzel, E. M., "Technical aspects of a demonstration tape for three-dimensional sound displays," *NASA TM102826*, 1990.

Begault, D. R., "Challenges to the successful implementation of 3-D sound," *Journal of the Audio Engineering Society*, Vol. 39, pp. 864-870, 1991.

Begault, D. R., "Preferred sound intensity increase for a sensation of half distance," *Perceptual and Motor Skills*, Vol. 72, pp. 1019-1029, 1991.

Begault, D. R., "Audio Spatialization Device for Radio Communications," *Report No. ARC 12013-1CU*, NASA-Ames Research Center, 1992.

Begault, D. R., "Binaural Auralization and Perceptual Veridicality," in *Audio Engineering Society 93rd Convention Preprint No. 3421 (M-3)*, Audio Engineering Society, New York, 1992.

Begault, D. R., "Perceptual effects of synthetic reverberation on three-dimensional audio systems," *Journal of the Audio Engineering Society*, Vol. 40, pp. 895-904, 1992.

Begault, D. R., "Perceptual similarity of measured and synthetic HRTF filtered speech stimuli," *Journal of the Acoustical Society of America*, Vol. 92, p. 2334, 1992.

Begault, D. R., "The Virtual Reality of 3-D Sound," in L. Jacobson (Ed.) *CyberArts: Exploring Art and Technology*, Miller-Freeman, San Francisco, CA, 1992.

Begault, D. R., "Call sign intelligibility improvement using a spatial auditory display," *Report No. 104014*, NASA-Ames Research Center, 1993.

Begault, D. R., "Head-up Auditory Displays for Traffic Collision Avoidance System Advisories: A Preliminary Investigation," *Human Factors*, Vol. 35, pp. 707-717, 1993.

Begault, D. R., *3-D Sound for Virtual Reality and Multimedia*, Academic Press Professional, Cambridge, MA, 1994.

Begault, D. R., and Erbe, T., "Multichannel spatial auditory display for speech communications," *Journal of the Audio Engineering Society*, Vol. 42, pp. 819-826, 1994.

Begault, D. R., "Virtual acoustic displays for teleconferencing: Intelligibility advantage for "telephone grade" audio," in *Audio Engineering Society 98th Convention (preprints)*, 1995.

Begault, D. R. and Pittman, M. T., "3-D Audio Versus Head Down TCAS Displays," *International Journal of Aviation Psychology*, (in press).

Begault, D. R. and Wenzel, E. M., "Technical aspects of a demonstration tape for three-dimensional auditory displays," *Report No. TM 102286*, NASA-Ames Research Center, 1990.

Begault, D. R. and Wenzel, E. M., "Headphone localization of speech," *Human Factors*, Vol. 35, pp. 361-376, 1993.

Begault, D. R. and Wenzel, E. M., "Techniques and applications for binaural sound manipulation in human-machine interfaces," in *International Journal of Aviation Psychology*, Vol. 2, pp. 1-22, 1992.

von Bekesy, G., *Experiments in Hearing*, McGraw-Hill, New York, 1960.

Benade, A., *Fundamentals of Musical Acoustics*, Dover Publications, New York, 1976. (Wave physics fundamentals of music.)

Bishop, G., ... Wenzel, E. M., et al., "Research Directions in Virtual Environments: Report of an NSF Invitational Workshop," *Computer Graphics*, Vol. 26, pp. 154-177, 1992.

Blauert, J., "An introduction to binaural technology," in R. Gilkey and T. Anderson (Eds.) *Binaural and Spatial Hearing*, Lawrence Elbaum Associates, Hillsdale, NJ, (in press). (A recent survey of the uses for 3D audio.)

Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, Cambridge, MA, 1983. (This is the standard book on the psychoacoustics of spatial hearing. Detailed and through.)

Blauert, J. (guest Ed.), "Special issue on auditory virtual environment and telepresence," *Applied Acoustics*, Vol. 36, Elsevier Applied Science, England, 1992.

Bloom, P. J., "Creating source elevation illusions by spectral manipulation," *Journal of the Audio Engineering Society*, Vol. 25, pp. 560-565, 1977.

Bregman, A. S., *Auditory Scene Analysis*, MIT Press, Cambridge, MA, 1990.

Bronkhorst, A. W. and Plomp, R., "The Effect of Head-Induced Interaural Time and Level Differences on Speech Intelligibility in Noise," *Journal of the Acoustical Society of America*, Vol. 83, pp. 1508-1516, 1988.

Burger, J. F., "Front-back discrimination of the hearing system," *Acustica*, Vol. 8, pp. 301-302, 1958.

Butler, R. A. and Belendiuk, K., "Spectral cues utilized in the localization of sound in the median sagittal plane," *Journal of the Acoustical Society of America*, Vol. 61, pp. 1264-1269, 1977.

Calhoun, G. L., Valencia, G. and Furness, T. A. III, "Three-dimensional auditory cue simulation for crew station design/evaluation," in *Proceedings of the Human Factors Society*, Vol. 31, pp. 1398-1402, 1987.

Cannon, R., *Dynamics of Physical Systems*, McGraw-Hill, New York, 1967. (Wave physics.)

Chan, C. J., "Sound Localization and Spatial Enhancement with the Roland Sound Space Processor," in *CyberArts: Exploring Art and Technology*, L. Jacobson (Ed.), pp. 95-104, Miller-Freeman Inc., San Francisco, CA, 1992.

Cherry, E. C., "Some Experiments of the Recognition of Speech with One and Two Ears," *Journal of the Acoustical Society of America*, Vol. 22, pp. 61-62, 1953.

Cherry, E. C. and Taylor, W. K., "Some further experiments on the recognition of speech with one and with two ears," *Journal of the Acoustical Society of America*, Vol. 26, pp. 549-554, 1954.

Churchland, P. S., *The Computational Brain*, MIT Press, Cambridge, MA, 1992.

Cohen, M. and Wenzel, E. M., "The Design of Multichannel Sound Interfaces," in W. Barfield and T. Furness III (Eds.) *Virtual Environments and Advanced Interface Design*, Oxford University Press, (in press).

Coleman, P. D., "Failure to localize the source distance of an unfamiliar sound," *Journal of the Acoustical Society of America*, Vol. 34(3), pp. 345-346, 1962.

Coleman, P. D., "An analysis of cues to auditory depth perception in free space," *Psychological Bulletin*, Vol. 60, pp. 302-315, 1963.

Critchley, M. and Henson, R. (Eds.), *Music and the Brain: Studies in the Neurology of Music*, Charles C. Thomas Publisher, Springfield, Illinois, 1977. (A collection of many engaging articles of interest to musicians.)

Doll, T. J., Gerth, J. M., Engleman, W. R. and Folds, D. J., "Development of simulated directional audio for cockpit applications," *USAF Report No. AAMRL-TR-86-014*, 1986.

Deutsch, Diana (Ed.), *The Psychology of Music*, Academic Press, 1982.

Dowling, W. J. and Harwood, D. L., *Music Cognition*, Academic Press, New York, 1986.

Durlach, N. I. and Colburn, H. S., "Binaural Phenomena," in E.C. Carterette and M. P. Friedman (Eds.) *Handbook of Perception*, Vol 4., New York, Academic Press, 1978.

Durlach, N. I., Rigopulos, A., Pang, X. D., Woods, W. S., Kulkarni, A., Colburen, H. S. and Wenzel, E. M., "On the externalization of auditory images," *Presence*, Vol. 1 (2), pp. 251-257, Spring 1992. (Aimed at virtual reality applications. See also Wenzel.)

Durlach, N. I. and Mavor, A. S., Eds., *Virtual Reality: Scientific and Technological Challenges: Report of the Committee on Virtual Reality Research and Development*, Washington, D.C., National Academy Press, 1994.

Fisher, H. and Freeman, S. J., "The role of the pinna in auditory localization," *Journal of Audio Research*, Vol. 8, pp. 15-26, 1968.

Fisher, S. S., Wenzel, E. M., Coler, C. and McGreevy, M. W., "Virtual interface environment workstations," in *Proceedings of the Human Factors Society*, Vol. 32, pp. 91-95, 1988.

Foster, S. H., *Convolvotron™ User's Manual*, Crystal River Engineering, Inc., 12350 Wards Ferry Road, Groveland, CA 95321, 1988.

Foster, S. H., Wenzel, E. M. and Taylor, R. M., "Real-time synthesis of complex acoustic environments [Summary]," in *Proceedings of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio & Acoustics*, New Paltz, New York, 1991.

Foster, S. H. and Wenzel, E. M., "Virtual acoustic environments: The Convolvotron [Summary]," *Computer Graphics*, Vol. 25 (4), p. 386, Demonstration system at the 1st annual "Tomorrow's Realities Gallery", SIGGRAPH '91, 18th ACM Conference on Computer Graphics and Interactive Techniques, Las Vegas, Nevada, July 27 - August 2, 1991.

Gardner, M. B. and Gardner, R. S., "Problem of Localization in the Median Plane: Effect of Pinnae Cavity Occlusion," *Journal of the Acoustical Society of America*, Vol. 53, pp. 400-408, 1973.

Gehring, B., *Focal Point™ 3D Sound User's Manual*, Gehring Research Corporation, 189 Madison Avenue, Toronto, Canada, M5R 2S6, 1990.

Gierlich, H. W., "The Application of Binaural Technology," *Applied Acoustics*, Vol. 36, pp. 219-244, 1992.

Gilkey, R. and Anderson, T., (Eds.), *Binaural and Spatial Hearing*, Lawrence Erlbaum Associates, Inc., New Jersey, (in press).

Hahn, J., "An Integrated Virtual Environment System," *Presence*, Vol. 2, pp. 353-360, 1944.

Hall, D., *Musical Acoustics*, 2nd Ed., Brooks/Cole Publishing, Belmont CA, 1991. (Wave physics fundamentals of music.)

Hartman, W. M., "Localization of sound in rooms," *Journal of the Acoustical Society of America*, Vol. 74, pp. 1380-1391, 1983.

HEAD Acoustics, *Binaural Mixing Console* [product literature], Contact: Sonic Perceptions, 114A Washington Street, Norwalk, CT 06854.

Helmholtz, H., *Sensations of Tone*, Dover Publications, New York, 1954. (A classic. Originally published in 1885. Still a very good book. Great experimental techniques.)

Humanski, R. A. and Butler, R. A., "The contribution of the near and far ear toward localization of sound sources on the median plan," *Journal of the Acoustical Society of America*, Vol. 83, pp. 2300-2310, 1988.

Kendall, G. S. and Martens, W. L., "Simulating the Cues of Spatial Hearing in Natural Environments," in *Proceedings of the International Computer Music Conference*, 1984.

Klayman, A. I., "SRS: Surround sound with only two speakers," *Audio*, Vol. 8, pp. 32-37, August 1992. (Probably the most successful commercial system for spatialized audio.)

Kistler, D. K. and Wightman, F. L., "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *Journal of the Acoustical Society of America*, Vol. 91, pp. 1637-1647, 1992.

Kramer, G. (Ed.), "Auditory Display: Sonification, Audification, and Auditory Interfaces," in *Proceedings Volume XVIII*, Santa Fe Institute Studies in the Sciences of Complexity, Reading MA, Addison-Wesley, 1994.

Kuhn, G. F., "Model for the interaural time differences in the azimuthal plane," *Journal of the Acoustical Society of America*, Vol. 62, pp. 157-167, 1977.

Loomis, J. M., Hebert, C. and Cicinelli, J. G., "Active localization of virtual sounds," *Journal of the Acoustical Society of America*, Vol. 88, pp. 1757-1764, 1990.

Macpherson, E. A., "On the role of head-related transfer function spectral notches in the judgement of sound source elevation," In G. Kramer (Ed.) *Proceedings of the 1994 International Conference on Auditory Displays*, Santa Fe, NM, (in press).

Makous J. C. and Middlebrooks, J. C., "Two-dimensional sound localization by human listeners," *Journal of the Acoustical Society of America*, Vol. 87, pp. 2188-2200, 1990.

McKinley, R. L. and Ericson, M. A., "Digital synthesis of binaural auditory localization azimuth cues using headphones," *Journal of the Acoustical Society of America*, Vol. 83, S18, 1988.

Mehrgardt, S. and Mellert, V., "Transformation characteristics of the external human ear," *Journal of the Acoustical Society of America*, Vol. 61, pp. 1567-1576, 1977.

Mershon, D. H. and King, L. E., "Intensity and reverberation as factors in the auditory perception of egocentric distance," *Perception and Psychophysics*, Vol. 18, pp. 409-415, 1975.

Middlebrooks, J. C., Makous, J. C. and Green, D. M., "Directional sensitivity of sound-pressure levels in the human ear canal," *Journal of the Acoustical Society of America*, Vol. 86, pp. 89-108, 1989.

Middlebrooks, J. C., and Green, D. M., "Directional dependence of interaural envelope delays," *Journal of the Acoustical Society of America*, Vol. 87, pp. 2149-2162, 1990.

Middlebrooks, J. C., "Narrow-band sound localization related to external ear acoustics," *Journal of the Acoustical Society of America*, Vol. 92, pp. 2607-2624, 1992.

Middlebrooks, J. C. and Green, D. M., "Sound Localization by Human Listeners," *Annual Review of Psychology*, Vol. 42, pp. 135-159, 1991.

Mills, A. W., "Auditory Localization", in J. V. Tobias (Ed.) *Foundations of Modern Auditory Theory*, Vol. II, pp. 303-348, Academic Press, New York, 1972. (A slightly dated but easy to understand survey.)

Mowbray, G. H. and Gebhard, J. W., "Man's senses as informational channels," in H. W. Sinaiko (Ed.) *Human Factors in the Design and Use of Control Systems*, pp. 115-149, Dover Publications, New York, 1961.

Oldfield, S. R. and Parker, S. P. A., "Acuity of sound localization: a topography of auditory space. I. Normal hearing conditions," *Perception*, Vol. 13, pp. 601-617, 1984.

Oldfield, S. R. and Parker, S. P. A., "Acuity of sound localization: a topography of auditory space. II. Pinna cues absent," *Perception*, Vol. 13, pp. 601-617, 1984.

Oldfield, S. R. and Parker, S. P. A., "Acuity of sound localization: a topography of auditory space. III. Monaural hearing conditions," *Perception*, Vol. 15, pp. 67-81, 1986.

Perrot, D. R., "Studies in the perception of auditory motion," in R. W. Gatehouse (Ed.) *Localization of Sound: Theory and Applications*, pp. 169-193, Amphora Press, Groton, CN, 1982.

Perrott, D. R., "Concurrent minimum audible angle: a re-examination of the concept of auditory spatial acuity," *Journal of the Acoustical Society of America*, Vol. 75, pp. 1201-1206, 1984.

Perrott, D. R., "Discrimination of the spatial distribution of concurrently active sound sources: Some experiments with stereophonic arrays," *Journal of the Acoustical Society of America*, Vol. 76, pp. 1704-1712, 1984.

Perrott, D. R. and Tucker, J., "Minimum audible movement angle as a function of signal frequency and the velocity of the source," *Journal of the Acoustical Society of America*, Vol. 83, pp. 1522-1527, 1988.

Perrott, D. R., Sadralodabai, T., Saberi, K. and Strybel, T. Z., "Aurally aided visual search in the central vision field: Effects of visual load and visual enhancement of the target," *Human Factors*, Vol. 33, pp. 389-400, 1991.

Persterer, A., "A very high performance digital audio processing system," in *Proceedings of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio & Acoustics*, New Paltz, New York, 1989.

Pierce, J. R., *The Science of Musical Sound*, revised edition, W. H. Freeman, New York, 1992. (Wave physics fundamentals of music.)

"The Physics of Music," *Scientific American*, W. H. Freeman and Company, San Francisco, CA, 1978. (Wave physics fundamentals of music.)

Plenge, G., "On the difference between localization and lateralization," *Journal of the Acoustical Society of America*, Vol. 56, pp. 944-951, 1974.

Plomp, R., *Aspects of the Tone Sensation*, Academic Press, London, 1976. (A compendium of psychoacoustic experiments and results.)

Lord Rayleigh [Strutt, J. W.], "On Our Perception of Sound Direction," *Philosophical Magazine*, Vol. 13, pp. 214-232, 1907.

Roffler, S. K. and Butler, R. A., "Factors that influence the localization of sound in the vertical plane," *Journal of the Acoustical Society of America*, Vol. 43, pp. 1255-1259, 1968.

Roffler, S. K., and Butler, R. A., "Localization of tonal stimuli in the vertical plane," *Journal of the Acoustical Society of America*, Vol. 43, pp. 1260-1266, 1968.

Rossing, T., *The Science of Sound*, Addison-Wesley, Reading, MA, 1990. (The acoustics of musical instruments is covered in an organized manner.)

Sakamoto, N., Gotoh, T., and Kimura, Y., "On 'out-of-head localization' in headphone listening," *Journal of the Audio Engineering Society*, Vol. 24, pp. 710-716, 1976.

Schroeder, M. R., "Digital Simulation of Sound Transmission in Reverberant Spaces," *Journal of the Acoustical Society of America*, Vol. 47, pp. 424-431, 1970.

Schubert, E. D., *Hearing: Its Function and Dysfunction*, Springer-Verlag/Wien, New York, 1980. (Although out of date, it was the state of the art for the 1980's, and is presented in breadth and depth.)

Seashore, C., *Psychology of Music*, Dover Publications, New York, 1967. (Although not explicitly referenced, the principles of Gestalt theorists come into play.)

Shaw, E. A. G., "The External Ear," in W. D. Keidel and W. D. Neff (Eds.) *Handbook of Sensory Physiology, Vol. V/1, Auditory System*, pp. 455-490, Springer-Verlag, New York, 1974.

Shinn-Cunningham, B. G., Lehnert, H., Kramer, G., Wenzel, E. M. and Durlach, N. I., "Auditory Displays," in R. Gilkey and T. Anderson (Eds.) *Binaural Hearing*, Lawrence Erlbaum Associates Inc., New Jersey, (in press).

Spiegle, J. M. and Loomis, J. M., "Auditory distance perception by translating observers," in *Proceedings of the IEEE Symposium in Research Frontiers in Virtual Reality*, San Jose, CA, October 25-26, 1993.

Strum, R. and Kirk, D., *First Principals of Discrete Systems and Digital Signal Processing*, Addison-Wesley, Reading, MA, 1989. (A good book for understanding convolution and spectral analysis having many figures.)

Sunier, J., "Binaural overview: Ears where the mikes are. Part I," *Audio*, Vol. 73, pp. 75-84, November 1989. (Excellent survey of binaural recording practices.)

Sunier, J., "Binaural overview: Ears where the mikes are. Part II," *Audio*, Vol. 73, pp. 49-57, December 1989. (Excellent survey of binaural recording practices.)

Takala, T., Hahn, J., Gritz, L., Geigel, J. and Lee, J., "Using physically-based models and genetic algorithms for functional composition of sound signals, synchronized to animated motion," *International Computer Music Conference (ICMC)*, Tokyo, Japan, September 10-15, 1993.

Thurlow, W. R. and Runge, P. S., "Effects of induced head movements on localization of direction of sound sources," *Journal of the Acoustical Society of America*, Vol. 42, pp. 480-488, 1967.

Wallach, H., "On sound localization," *Journal of the Acoustical Society of America*, Vol. 10, pp. 270-274, 1939.

Wallach, H., *"The role of head movements and vestibular and visual cues in sound localization,"* Journal of Experimental Psychology*, Vol. 27, pp. 339-368, 1940.

Warren, D. H., Welch, R. B. and McCarthy, T. J., "The Role of Visual-Auditory 'Compellingness' in the Ventriloquism Effect: Implications for Transitivity Among the Spatial Senses," *Perception and Psychophysics*, Vol. 30, pp. 557-564, 1981.

Watkins, A. J., "Psychoacoustical aspects of synthesized vertical locale cues," *Journal of the Acoustical Society of America*, Vol. 63, pp. 1152-1165, 1978.

Welch, R. B., *Perceptual Modification: Adapting to Altered Sensory Environments*, New York, Academic Press, 1978.

Wenzel, E. M., "Perceptual factors in virtual acoustic displays" [Invited Keynote Speaker], in *Proceedings of ICAT'94*, 4th International Conference on Artificial Reality and Tele-Existence, Tokyo, Japan, pp. 83-98, 1994.

Wenzel, E. M., "Spatial Sound and Sonification," in G. Kramer (Ed.) *Auditory Display: Sonification, Audification, and Auditory Interfaces*, Addison-Wesley, Reading, MA, pp. 127-150, 1994.

Wenzel, E. M. and Foster, S. H., "Perceptual consequences of interpolating head-related transfer functions during spatial synthesis," in *Proceedings of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio & Acoustics*, New Paltz, New York, October 17-20, 1993.

Wenzel, E. M., Gaver, W., Foster, S. H., Levkowitz, H. and Powell, R., "Perceptual vs. hardware performance in advanced acoustic interface design," in *Proceedings of INTERCHI'93*, Conference on Human Factors in Computing Systems, Amsterdam, pp. 363-366, 1993.

Wenzel, E. M., Arruda, M., Kistler, D. J. and Wightman, F. L., "Localization using non-individualized head-related transfer functions," *Journal of the Acoustical Society of America*, Vol. 94, pp. 111-123, 1993.

Wenzel, E. M., "Launching sounds into space," in L. Jacobson (Ed.) *CyberArts: Exploring Art and Technology*, Miller-Freeman Inc., San Francisco, CA, 1992.

Wenzel, E. M., "Three-dimensional virtual acoustic displays," in M. Blattner and R. Dannenberg (Eds.) *Multimedia Interface Design*, ACM Press, New York, 1992.

Wenzel, E. M., and Foster, S. H., *"Virtual Acoustic Environments. [Summary: demonstration system],"* in *Proceedings of the CHI'92*, ACM Conference on Computer-Human Interaction, Monterey, CA, p. 676, 1992.

Wenzel, E. M., "Localization on virtual acoustic displays," *Presence*, Vol. 1, pp. 80-107, Winter 1992.

116

Wenzel, E. M., "Virtual Acoustic Displays: Localization in Synthetic Acoustic Environments [Plenary speech]," in *Proceedings of Speech Tech '92*, February 4-5, New York, NY, 1992.

Wenzel, E. M., "Three-dimensional virtual acoustic displays," *NASA TM103835*, 1991.

Wenzel, E. M., Wightman, F. L. and Kistler, D. J., "Localization of non-individualized virtual acoustic display cues," in *Proceedings of the CHI'91*, ACM Conference on Computer-Human Interaction, New Orleans, LA, April 27-May 2, 1991.

Wenzel, E. M., Stone, P. K., Fisher, S. S. and Foster, S. H., "A system for three-dimensional acoustic 'visualization' in a virtual environment workstation," in *Proceedings of the IEEE Visualization '90 Conference,* San Francisco, CA October 23-26, pp. 329-337, 1990.

Wenzel, E. M., *"Virtual acoustic displays,"* in *Human Machine Interfaces for Teleoperators and Virtual Environments*, Santa Barbara, CA, March 4-9, NASA Conference Publication 10071, 1990.

Wenzel, E. M., and Foster, S. H., "Real-time digital synthesis of virtual acoustic environments," *Computer Graphics*, 1990.

Wenzel, E. M., Foster, S. H., Wightman, F. L. and Kistler, D. J., "Real-time Digital Synthesis of Localized Auditory Cues Over Headphones," in *Proceedings of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio & Acoustics*, New Paltz, NY, October 15-18, 1989.

Wenzel, E. M., Foster, S. H., Wightman, F. L. and Kistler, D. J., "Real-time Synthesis of Localized Auditory Cues," in *Proceedings of CHI'89*, ACM Conference of Computer-Human Interaction, Austin, TX, April 30 - May 5, 1989.

Wenzel, E. M., Wightman, F. L., Kistler, D. J. and Foster, S. H., "Acoustic origins of individual differences in sound localization behavior," *Journal of the Acoustical Society of America*, Vol. 84, S79(A), 1988.

Wenzel, E. M., Wightman, F. L., and Foster, S. H., "A virtual display system for conveying three-dimensional acoustic information," in *Proceedings of the Human Factors Society*, Vol. 32, pp. 86-90, 1988.

Wenzel, E. M., Fisher, S. S., Wightman, F. L. and Foster, S. H., "Application of auditory spatial information in virtual display systems," *CHABA Symposium on Sound Localization*, Sponsored by the National Academy of Science and the AFOSR, Washington, D. C., October 14-16, 1988.

Wenzel, E. M., Wightman, F. L. and Foster, S. H., "Development of a three-dimensional auditory display system," *SIGCHI Bulletin*, Vol. 20, pp. 52-57, 1988.

Wenzel, E. M., Wightman, F. L. and Foster, S. H., "Development of a three-dimensional auditory display system," in *Proceedings of CHI'88*, ACM Conference on Computer-Human Interaction, Washington, D. C., May 15-19, 1988.

Wightman, F. L. and Kistler, D. J., "Headphone Simulation of Free-field Listening I: Stimulus Synthesis," *Journal of the Acoustical Society of America*, Vol. 85, pp. 858-867, 1989.

Wightman, F. L. and Kistler, D. J., "Headphone Simulation of Free-field Listening II: Psychophysical Validation," *Journal of the Acoustical Society of America*, Vol. 85, pp. 868-878, 1989.

Wightman, F. L. and Kistler, D. J., "The Dominant Role of Low-frequency Interaural Time Differences in Sound Localization," *Journal of the Acoustical Society of America*, Vol. 91, pp. 1648-1661, 1992.

Wightman, F. L. and Kistler, D. J., "Multidimensional Scaling Analysis of Head-Related Transfer Functions," in *Proceedings of the ASSP (IEEE) Workshop on Applications of Signal Processing to Audio and Acoustics*, IEEE Press, New York, 1993.

Wightman, F. L., Kistler, D. J. and Anderson, K., "Reassessment of the role of Head Movements in Human Sound Localization," *Journal of the Acoustical Society of America*, Vol. 95, pp. 3003-3004, 1994.

Wightman, F. L. and Kistler, D. J., "The Importance of Head Movements for Localizing Virtual Auditory Display Objects," in G. Kramer (Ed.) *Proceedings of the 1994 International Conference on Auditory Displays*, Santa Fe, NM, (in press).

Zahorik, P. A., Kistler, D. J. and Wightman, F. L., "Sound Localization in varying virtual acoustic environments," in G. Kramer (Ed.) *Proceedings of the 1994 International Conference on Auditory Displays*, Santa Fe, NM, (in press).

Zurek, P. M., "Binaural Advantages and Directional Effects in Speech Intelligibility," in G. A. Studebaker and I. Hochberg (Eds.) *Acoustical Factors Affecting Hearing Aid Performance*, Allyn and Bacon, Needham Heights, MASS, 1993.

APPENDIX A: LIST OF DEFFINITIONS AND ABBREVIATIONS

λ	wavelength
f	frequency
c	speed of light.
2D	two dimension
3D	three dimension
AD	Analog-to-Digital
annabelle	name of the workstation which runs NPSNET-3DSS
C++	A Programming Language
CCRMA	Center for Computer Research in Music and Acoustics
CD	Compact Disc (16 bit audio)
CP-1 Plus	Lexicon Digital Audio Environment Processor
CPU	Central Processing Unit
DAT	Digital Audio Tape
dB	Decibel
DA	Digital-to-Analog
DIN	Deutsche Industri Norm
DIS	Distributed Interactive Simulation
DSP	Digital Signal Processor/Processing
EMAX II	16 bit digital sound system keyboard/sampler manufactured by E-Mu Corporation [EMU89]
Ensoniq DP/4	MIDI capable parallel effects processor containing 4 processors manufactured by Ensoniq Corporation [ENSO92a]
FIR	Finite Impulse Response

HF	High Frequency
HRTF	Head-Related Transfer Function
IEEE	Institute of Electrical and Electronics Engineers
IID	Interaural Intensity Difference
IRCAM	Institute of Research and Coordination of Acoustics and Music
Iris Indigo	Silicon Graphics Workstation
ITD	Interaural Time Difference
IP	Internet Protocol
JASA	Journal of the Acoustical Society of America
LAN	Local Area Network
MAC	abbreviation for an Apple Macintosh Computer
MHz	Mega Hertz
MIDI	Musical Instrument Digital Interface
ms	milliseconds
NPS	Naval Postgraduate School
NPSNET	Naval Postgraduate School Networked Vehicle Simulator
NPSNET-PAS	NPSNET-Polyphonic Audio Spatializer
NRG	NPSNET Research Group
PE	Precedence Effect
PDU	Protocol Data Unit
Polhemus Fastrack	Motion Tracker
SC	Sound Cube
SCM	Sound Cube Model
SE	Synthetic Environment

SR	Synthetic Reverberation
SGI	Silicon Graphics Incorporated
Speed of Sound	335.28 meters per second in air at sea level and 70 degrees Fahrenheit
RAM	Random Access Memory
VE	Virtual Environment
VLF	Very Low Frequency
ZIPI	name of new language/protocol for describing music which makes improvements on MIDI

APPENDIX B: NPSNET-3DSS SETUP GUIDE

A.   HARDWARE SETUP

The following items are required to be in the defined position or setup configuration before starting NPSNET-3DSS.

STEP 1 - SCSI Removable Hard Drive - This is the SCSI hard drive that is attached to the EMAX II. This drive must be turned on before the EMAX II. The on/off switch is located in the upper right hand corner of the rear panel. When facing the front of the drive, this would be on the left side. Once this drive is turned on, the yellow lights on the front panel will begin blinking. When the drives have successfully booted, the green lights will be lit and the yellow light extinguished. This operation takes approximately 20 seconds.

STEP 2 - EMAX II Sampler - Move the slider marked *VOLUME* to the lowest position possible. Facing the front of the EMAX II, the on/off switch is located on the back panel to the right. Turn this switch on and allow approximately 25 seconds for the EMAX II to boot up. Once booted, press the button marked *SETUP*. The LED readout will show the words *Sequencer Setup* in the top half of the window. Next, press the numeral *6* on the EMAX numeral keypad located just below the LED readout. The LED should now display the words *Super Mode: off* in the top half of the window. Now, press the button marked *ON YES* located to the left of the EMAX numeric keypad in order to select yes. Now, the display in the upper half of the LED window should read *Super Mode: on*. Next, press the button marked *ENTER* located to the right rear of the numeric keypad. Next, press the *SETUP* button located up and to the right of the *ENTER* button. The LED display should now show *P00 Untitled* in the upper half of the window.

STEP 3 - Mixing Console - On the Allen & Heath GL2 mixing console ensure all volume sliders are set at the bottom. There is no on/off switch, for the mixing console is always on. However, to ensure that the mixing console is on, there should be a green light illuminated which is located just above the headphones connector jack in the far upper right portion of the mixing console. The Allen & Heath mixer uses a dB scale for volume output.

This means that a position of 0 is full volume, a position above 0 is a dB boost, and a position below 0 is a dB reduction. Note, this does not refer to the physical position of the slider, but rather to the scale drawn on the console next to each slider. Move the white sliders for channels 1, 2, 3, 4, 5, 6, 7, and 8 so that the black line in the center of the slider lines up with the labeled white dot position markers. Move the yellow sliders for the master volume control labeled L and R so that the black line in the center of the sliders is lined up with the labeled white dot position markers. Ensure the pan pot settings for channels 1, 3, 5 and 7 are set to L: brown knob turned all the way to the left. Ensure the pan pot settings for channels 2, 4, 6, and 8 are set to R: brown knob turned all the way to the right. Ensure all push-button switches are set in their proper positions as indicated by the white dot position markers.

STEP 4 - Ensoniq DP/4 - There are two of these signal processors located in the top two spaces of the audio rack. Press the on/off switch, which is located on the right foremost position on the front panel, for each unit to the on position. The DP/4's take approximately 5 seconds to boot. Ensure the volume settings for both the top and bottom DP/4s are set with channels 1, 2, 3, and 4 (top and bottom) at one notch mark past the halfway point.

STEP 5 - RAMSA Subwoofer Processor - Press the button marked *Power* in the middle of the front panel located just under the words *Studio 3*. A red light will illuminate to indicate power is on. Note, there is an additional power switch located to the far right of this front panel, however, this switch should not be turned on.

STEP 6 - Carver Power Amplifier - Press the button marked *Power*. Ensure the volume settings for each channel are at maximum volume. This is when the level controls marked *L* and *R* are rotated fully in the clockwise direction.

STEP 7 - RAMSA Power Amplifiers - These amplifiers are located in the bottom two spaces of the audio rack. Press the switch marked *Power* to the on position for both top and bottom RAMSA power amplifiers. Ensure that the volume is set to 50% for both the A and B channels of each of the two amplifiers. This will put the position indicators facing directly upward.

STEP 8 - Execute Program - The final step in bringing up the sound system is to start the software program. This procedure is detailed in the next section. Once the software is started, increase the slider marked *volume* on the EMAX II to the desired position. This slider will control the overall volume of the system. Use this slider to adjust overall volume up and down as desired, for it equally affects all subchannels on the EMAX II. Alteration of any of the other volume controls throughout the system will result in the speakers being thrown out of balance and will severely degrade the localization/spatialization capabilities of the system.

B.   SOFTWARE EXECUTION

The only machine that supports the NPSNET-3DSS is *annabelle*. So, you must first login or rxterm to *annabelle* before accessing the software. The NPSNET-3DSS software currently can be found in the following directory: */workd/storms/npsnet-midi-sound/demo-3d-research*. So, to run the sound server, you will need to change directory to this directory. The executable is titled *NPS3DSS*. However, simply typing this command at the prompt will not properly start the program. In order to increase modularity and to increase flexibility with loading multiple terrains, there is a series of switches/options that must be selected at run time. Furthermore, a script file called *demo-midi-sound* has been written incorporating these various switches. Thus, the simplest way to run the sound server is to type *demo-midi-sound* and hit return. You will next see on the screen the proper format to select the various terrains (i.e. benning, hunterliggett, trg, etc.).

1. Command Line Options

If you elect not to use the script file *demo-midi-sound*, you can customize the sound server for your particular application. To use the command line switches, type *NPS3DSS* followed by the desired switches. For example, *NPS3DSS -w* would run the sound server without the graphics window. The following is a list of possible command line switches. All of the switches do not need to be set. However, the amount and type of switches to use

depends on the particular NPSNET application that is to be run. This list can also be obtained by typing *NPS3DSS -h* at the command prompt.

-h	{for help}
-i \<interface\>	{to choose network interface}
-c or *-f* \<config file\>	{to read config file}
-e \<machine name\>	{to choose master machine}
-s \<site\>	{to choose master site}
-o \<host\>	{to choose master host}
-n \<entity\>	{to choose master entity}
-x \<exercise ID\>	{to set exercise ID}
-d	{to debug, no midi output}
-m	{to use Multicast}
-p \<port\>	{to set UDP port}
-g \<group\>	{to set Multicast group}
-t \<ttl\>	{to set Multicast ttl}
-w	{no graphics window}
-b \<bank num\>	{to select midi bank}
-v \<environment file\>	{to select env_snd.dat file}
-a	{to test sound directions}

2. Command Line Usage

-h: This simply outputs the list of possible switches to the screen.

-i: This specifies which ethernet interface to use. (There can be more than one per machine, however, all of our machines have exactly one. The name of the interface on the SGI Reality Engine equipped machines is *et0* and on all others is *ec0*.)

-c or *-f*: This switch allows for different configuration files to be read upon execution. Some of the configuration files available are: *config.trg*, *config.benning*, and *config.hl*. These configuration files contain the following data: the name of the master or

host machine, the specification of the round world coordinates that are to be used, the exercise ID number, the environmental data file, and the network file. If any of these parameters are given by another command line switch, the config file parameters will be overridden.

-e: This determines which machine will be defined as the host entity. This is important, as the host position will act as the center of the sound world and all sounds generated will be based on this entity's position. The default host is *meatloaf*. For example, *-e* gravy3, would make the user on gravy3 be the host.

-s: Use this switch to choose the master site.

-o: Use this switch to choose the master host.

-n: Use this switch to choose the master entity. This switch will set up the network portion of the program to read packets using a multicast wrapper around the data packets being sent. This allows NPSNET-3DSS and NPSNET to be used over the internet.

-x: This is the DIS simulation exercise identifier. It is required to allow the network code to read only the packets that apply to the selected exercise. This identifier must be obtained from the user that initiates the simulation exercise.

-d: This will disable the transmission of MIDI data to the sampler for purposes of debugging program changes.

-m: Use this switch to enable Multicast (as opposed to Broadcast).

-p: This is the network port number (UDP) which is required for multicast.

-g: This is the multicast group number which is required for multicast.

-t: This is the multicast ttl. This determines the length of time a packet will stay alive on the internet and how far it will reach. This is required for multicast.

-w: If run on a less capable machine this will prevent the graphic display window from being drawn. Note, MIDI data output is not affected.

-b: This determines the bank number that the EMAX II will load upon execution. The default is bank 8, which is standard for all terrains currently being used by NPSNET.

The switch is invoked with a bank number as an argument. Example, *-b 5*, would load bank 5 upon execution.

-v: This switch enables the loading of the environmental data file.This provides the capability to load different geographic data for various environmental sound effects. Each terrain has many different properties and the environmental data is completely different. For example, *-v* environ_snd.dat, will load this file of geographic points with their associated sound data.

-a: This will perform a self-test of the audio system by playing sounds in the individual speakers and in the following order: lower right front, lower left front, lower right back, lower left back, upper right front, upper left front, upper right back, and upper left back. If only using four speakers, the same test is performed, so the sounds are basically played in each speaker twice. This switch is provided for verifying setup when debugging changes to the program. If the sounds are heard in the correct order, the directional algorithm can be assumed to be working correctly. This switch is also very handy in verifying the external audio system when reconfiguring or resetting up the hardware. It is very common to cross audio channels when setting up the system.

APPENDIX C: HARDWARE WIRING DIAGRAMS

This appendix contains a more detailed description of the hardware wiring digrams for both the partial and full sound cube implementation. The wiring diagrams are identical for both partial and full sound cube implementation except as noted when routing the audio signals from the mixing board to the amplifiers/speakers.

A.   COMPUTER TO SAMPLER
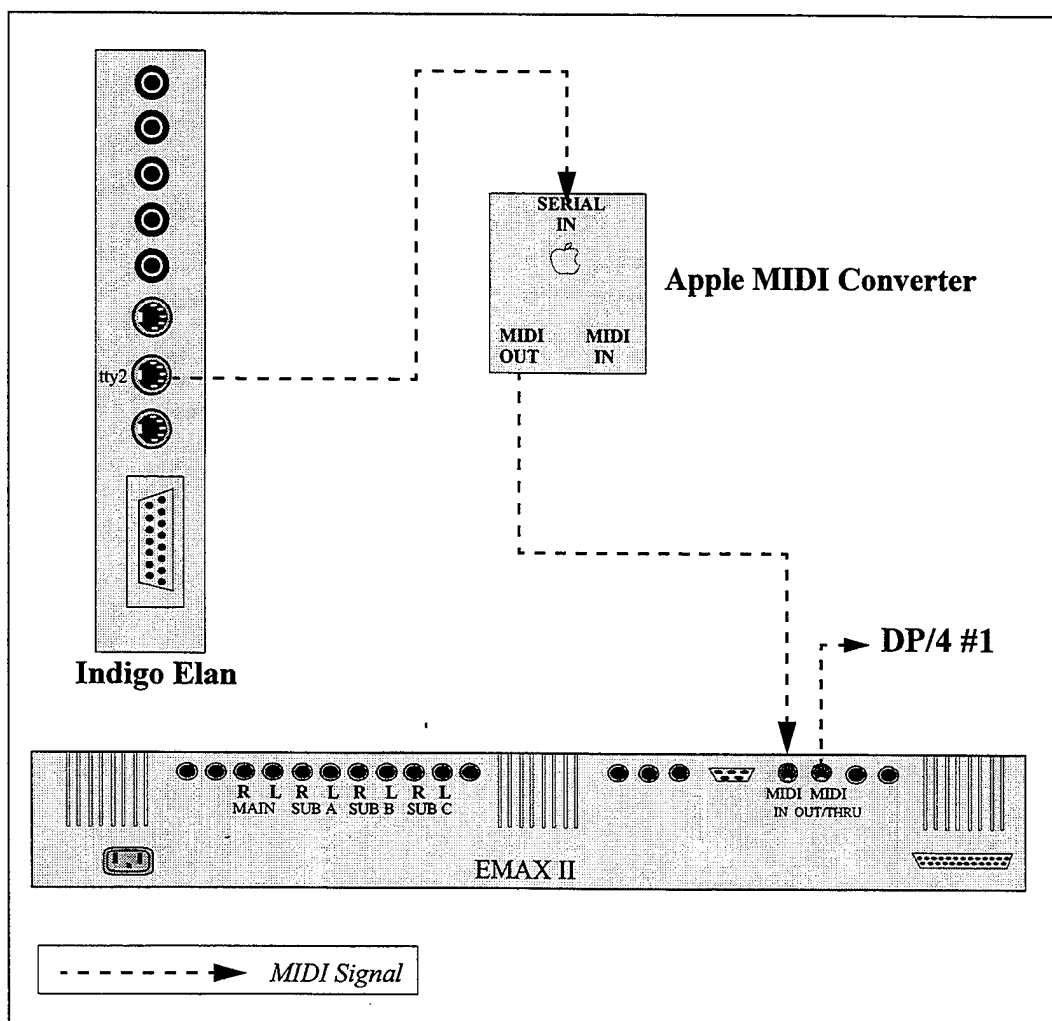
Figure 28: Computer to Sampler Wiring Diagram.

B. SAMPLER TO MIXING BOARD
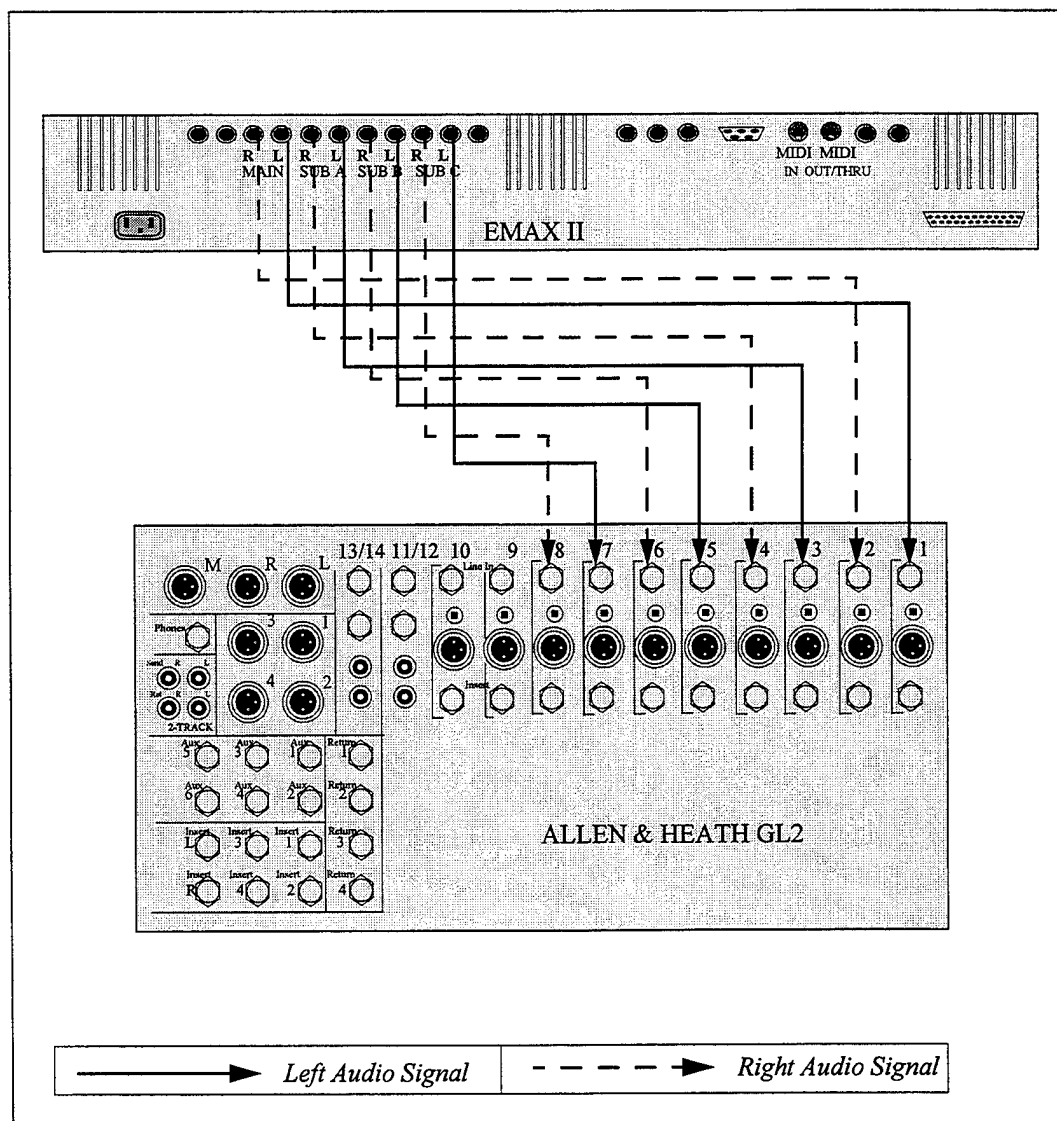
Figure 29: Sampler to Mixing Board Wiring Diagram.

C. MIXING BOARD TO DIGITAL SIGNAL PROCESSORS
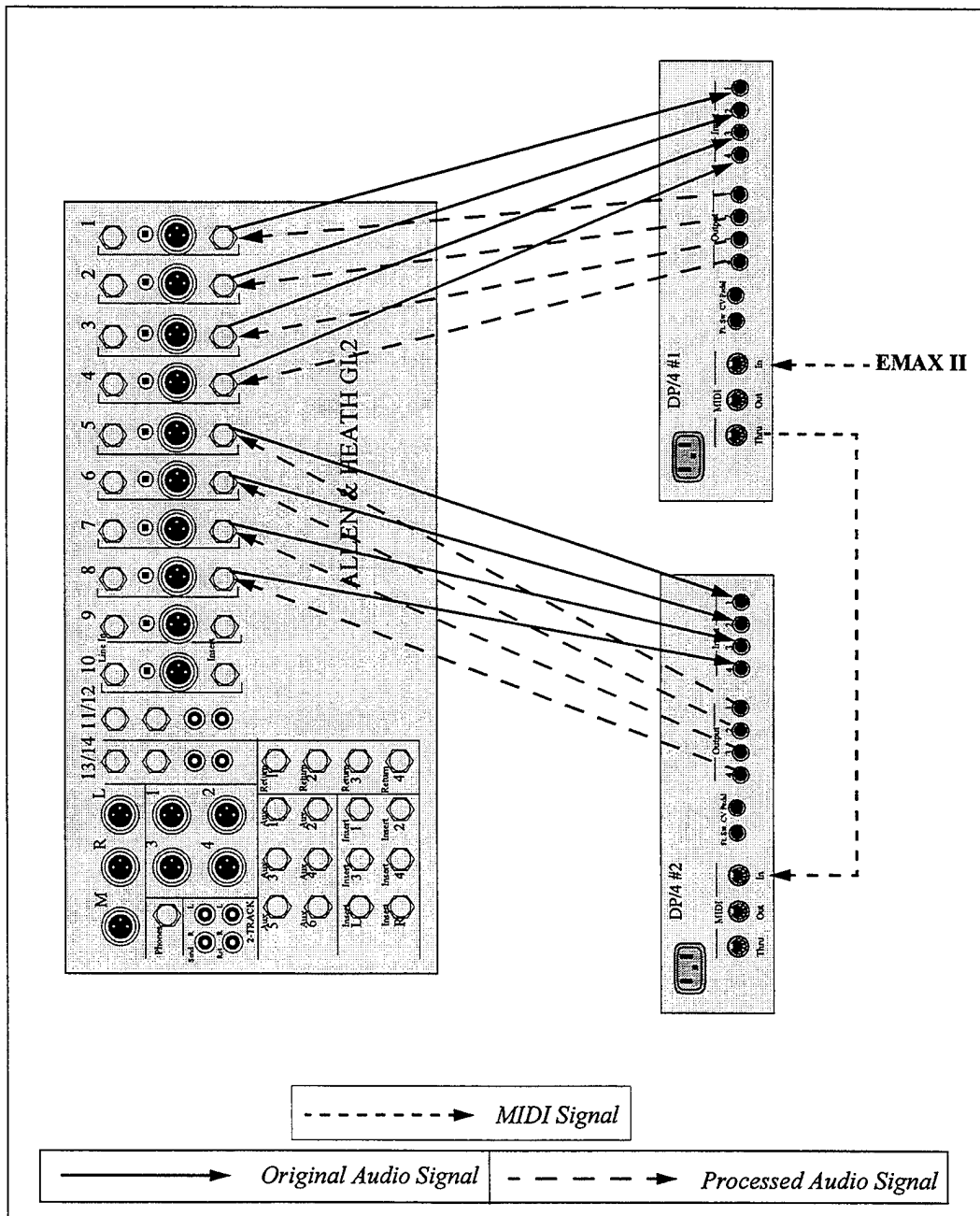
Figure 30: Mixing Board To DSPs Wiring Diagram.

D. MIXING BOARD TO AMPLFIERS/SPEAKERS
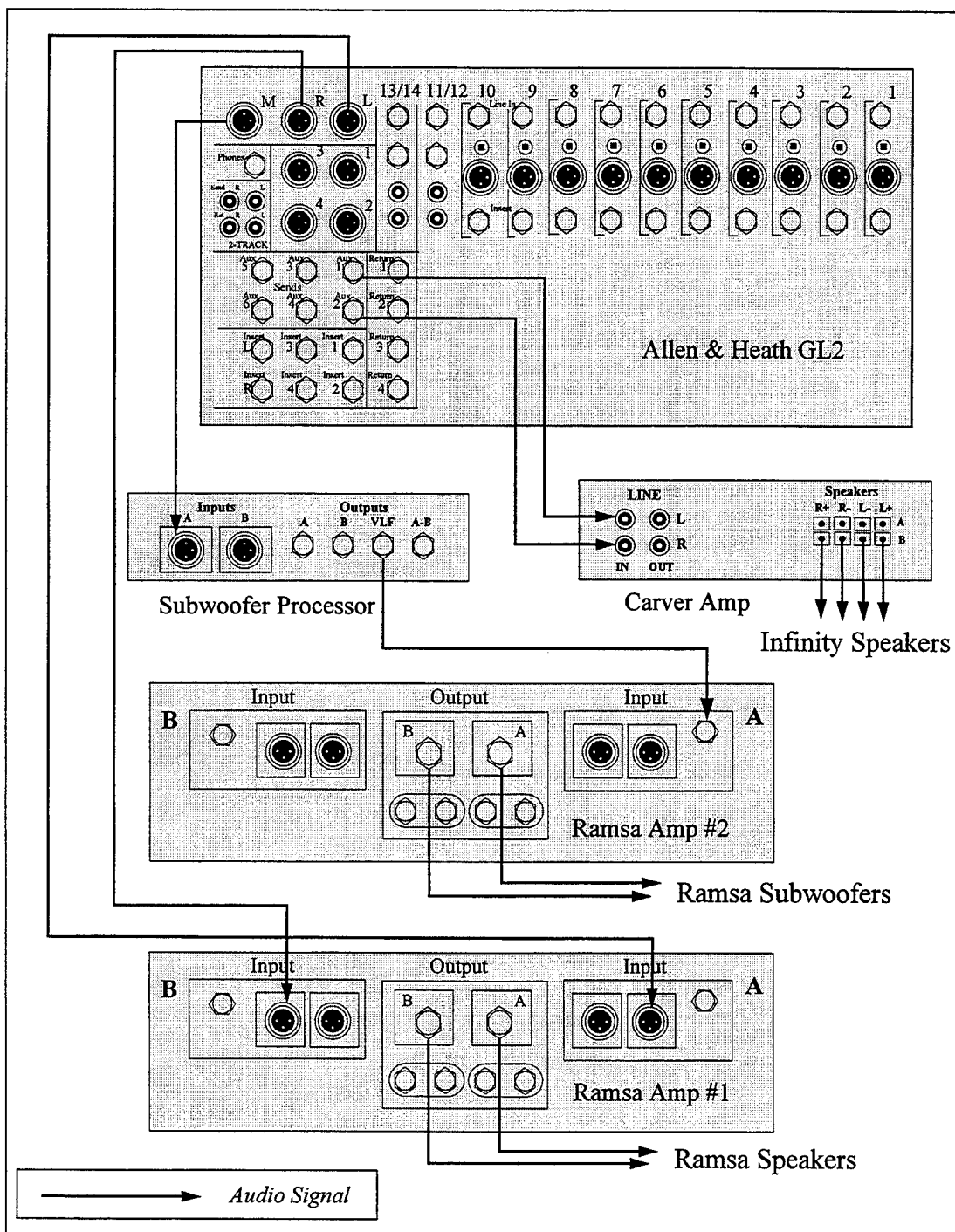
Figure 31: Partial SC Mixing Board to Amplifiers Wiring Diagram.
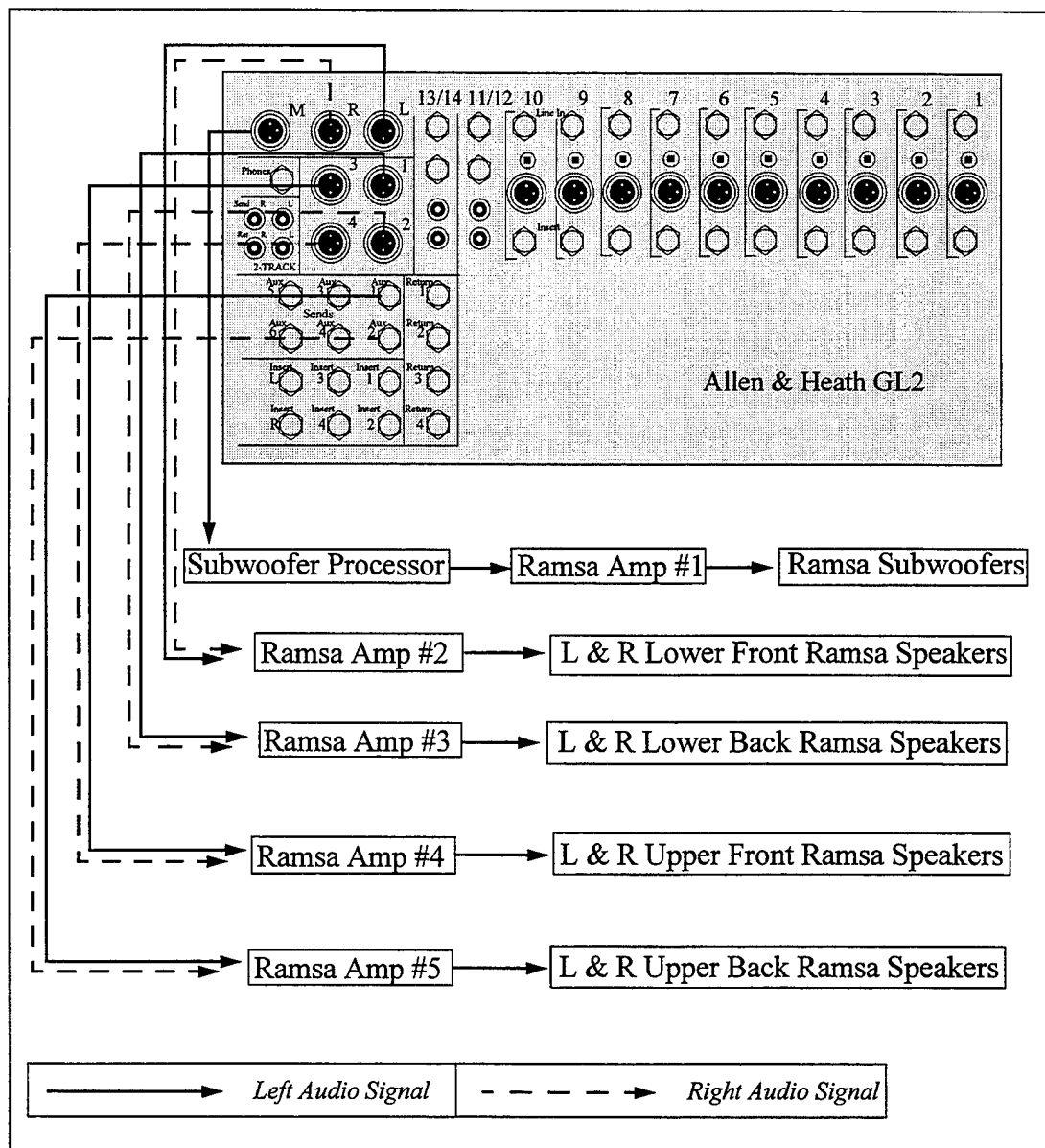
Figure 32: Full SC Mixing Board to Amplifiers Wiring Diagram.

APPENDIX D: EMAX II CONFIGURATION AND USE

This appendix serves as a guide to understanding how the EMAX II is configured for use with NPSNET-3DSS. For a detailed understanding of how to use the EMAX II consult the owner's manual (see [EMU89]). However, because the EMAX II has so much built-in functionality, the owner's manual alone is not very helpful. To better understand how the EMAX II is used in this research effort, one should look at both Dahl's and Roesli's Master's Thesis (see [DAHL92] and [ROES94]). Also, calling the technical assistants from E-mu Corporation, the makers of the EMAX II, can be very helpful. Nevertheless, the following are some key areas of interest that must be understood in order to gain an understanding as to how the EMAX II is configured and utilized in this research effort.

A.   SOUND BANK CONSTRUCTION

Besides MIDI, which is critical to know for understanding the EMAX II, the most fundamental concept is that of the sound bank. The sound bank is to the EMAX II what an operating system is to a computer. The sound bank determines which sounds can be played, how they should be played, where the sounds should be output, and how MIDI commands can access and manipulate the sounds. The sound bank for NPSNET-3DSS consists of sequences which are made up from individual presets. The presets usually contain discrete sounds while the sequences play continuous sounds.

Bank number eight, named *3DSnd NPSNET*, is the current sound bank used for NPSNET-3DSS. This bank is configured with four sequences. Sequence *01 Theme* contains a musical arrangement that is played when there are no hosts on the network. Sequence *02 Activated* contains a voice message that says the NPSNET sound server is activated. Sequence *03 Deactiv* contains a voice message that says the NPSNET sound server is deactivated. These sequences were written by John Roesli, the creator of NPSNET-PAS and have remained unchanged for use in NPSNET-3DSS. Incidently, it is Roesli's voice that is used for the activated and deactivated messages. Sequence *00 SFX* is

the heart of NPSNET-3DSS containing all the possible sounds that can be played corresponding to sound events in the VE of NPSNET. This sequence was originally configured for use in NPSNET-PAS, but has now been reconfigured for use in NPSNET-3DSS. The bulk of the reconfiguration lies in the presets, for it is the preset which determines the output port on the EMAX II. There are eight user selectable audio output ports on the EMAX II: MAIN R, MAIN L, Sub A R, Sub A L, Sub B R, Sub B L, Sub C R, and Sub C L. There are two more audio output ports: Headphones and Mono Mix, but these ports merely sum the output of the Main R and Main L output ports. Figure 33 depicts the location of these output ports.
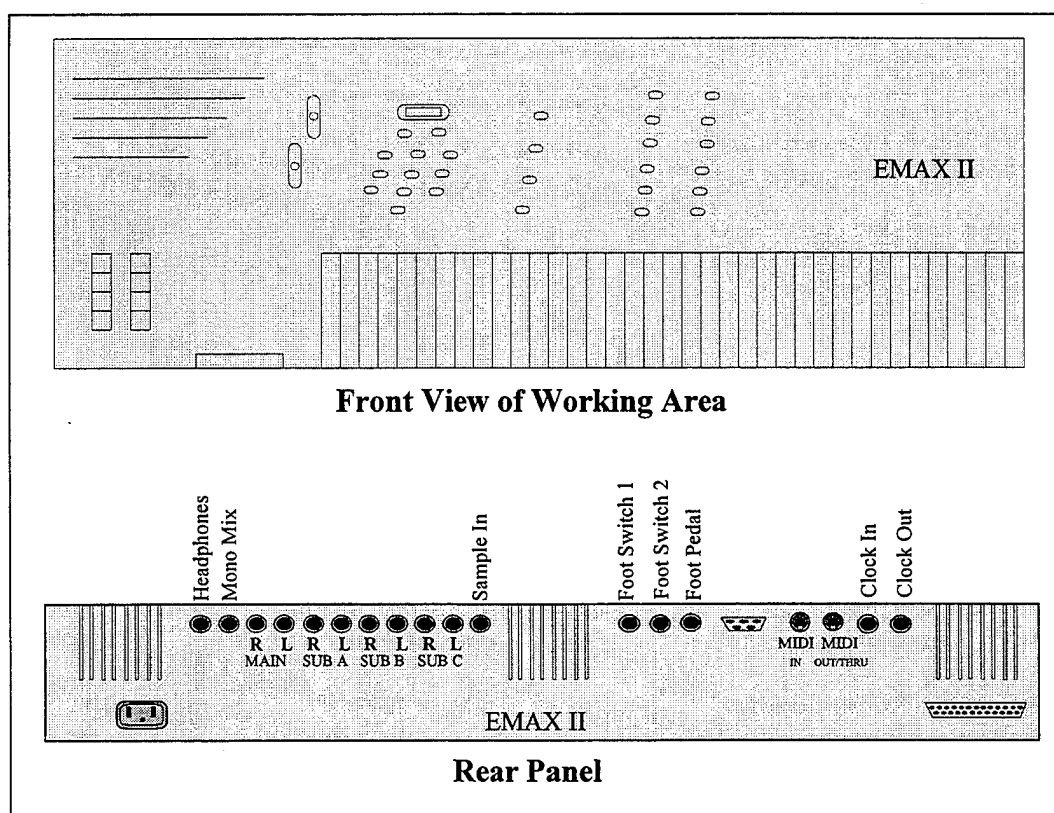
Figure 33: EMAX II Front View and Rear Panel.

In order to generate the eight independent sounds needed for each of the eight speakers of the SC, eight copies of the same preset (a particular sound) have been assigned to eight different outputs panned either to the left or to the right on the EMAX II. These presets make up the majority of the sequence *00 SFX*. The remainder of the sequence

contains the vehicle sounds. The presets are assigned individual MIDI channels to give MIDI commands direct access to the presets. The currently assigned MIDI channels are indicated in the tables that follow. Furthermore, the EMAX II itself can also be assigned a MIDI channel to distinguish the EMAX II from other daisy changed MIDI devices. Currently, the EMAX II has been assigned MIDI channel fifteen.

B.   SOUND BANK CONFIGURATION TABLES

In order to play a certain preset, it must be assigned a note value on the EMAX II. These note values are usually consistent among MIDI devices, but the EMAX II does not conform to typical note values as was discovered by Dahl in the development of NPSNET-Sound. The correct note values assigned to the EMAX II are listed in Table 1.

Octave	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
0	18	19	1A	1B	1C	1D	1E	1F	20	21	22	23
1	24	25	26	27	28	29	2A	2B	2C	2D	2E	2F
2	30	31	32	33	34	35	36	37	38	39	3A	3B
3	3C	3D	3E	3F	40	41	42	43	44	45	46	47
4	48	49	4A	4B	4C	4D	4E	4F	50	51	52	53
5	54	55	56	57	58	59	5A	5B	5C	5D	5E	5F
6	60	61	62	63	64	65	66	67	68	69	6A	6B

Table 1.   Hex Value Equivalents of EMAX II Keys. From [DAHL92].

Once given the proper note values, we can correctly setup the presets. The following tables of presets, which were originally setup by Roesli, have been changed to reflect the current configuration of the presets which define the majority of sequence *00 SFX*. In the tables, *Sample* refers to the type of sound sampled. *Note Value* refers to the particular note on the EMAX II that the sample has been assigned to in the sequence. *Hex Value* coincides with

the *Note Value* assignments and are used for the sending of MIDI commands to access the note.

Sample	Note Value	Hex Value	Output Channel	Pan Setting
Rifle	C-2	0x30	Main	Right
Rifle Large	D-2	0x32	Main	Right
Rile-Auto	E-2	0x34	Main	Right
M-60	F-2	0x35	Main	Right
25mm	G-2	0x37	Main	Right
Explosion1	A-2	0x39	Main	Right
Explosion2	B-2	0x3B	Main	Right
Explosion3	C-3	0x3C	Main	Right
Exposion4	D-3	0x3E	Main	Right
Explosion5	E-3	0x40	Main	Right
Explosion6	F-3	0x41	Main	Right
Sm. Missile	G-3	0x43	Main	Right
Med.Missile	A-3	0x45	Main	Right
Lg. Missile	B-3	0x47	Main	Right
Cannon1	C-4	0x48	Main	Right
Cannon2	D-4	0x4A	Main	Right
Lg.Artillery	E-4	0x4C	Main	Right
M1 Fire	F-4	0x4D	Main	Right
Seagulls	G-4	0x4F	Main	Right
Crickets	A-4	0x51	Main	Right

Table 2: Preset 01 (MIDI Channel 01).

Sample	Note Value	Hex Value	Output Channel	Pan Setting
Rifle	C-2	0x30	Main	Left
Rifle Large	D-2	0x32	Main	Left
Rile-Auto	E-2	0x34	Main	Left
M-60	F-2	0x35	Main	Left
25mm	G-2	0x37	Main	Left
Explosion1	A-2	0x39	Main	Left
Explosion2	B-2	0x3B	Main	Left
Explosion3	C-3	0x3C	Main	Left
Exposion4	D-3	0x3E	Main	Left
Explosion5	E-3	0x40	Main	Left
Explosion6	F-3	0x41	Main	Left
Sm. Missile	G-3	0x43	Main	Left
Med. Missile	A-3	0x45	Main	Left
Lg. Missile	B-3	0x47	Main	Left
Cannon1	C-4	0x48	Main	Left
Cannon2	D-4	0x4A	Main	Left
Lg.Artillery	E-4	0x4C	Main	Left
M1 Fire	F-4	0x4D	Main	Left
Seagulls	G-4	0x4F	Main	Left
Crickets	A-4	0x51	Main	Left

Table 3: Preset 02 (MIDI Channel 02).

Sample	Note Value	Hex Value	Output Channel	Pan Setting
Rifle	C-2	0x30	Sub A	Right
Rifle Large	D-2	0x32	Sub A	Right
Rile-Auto	E-2	0x34	Sub A	Right
M-60	F-2	0x35	Sub A	Right
25mm	G-2	0x37	Sub A	Right
Explosion1	A-2	0x39	Sub A	Right
Explosion2	B-2	0x3B	Sub A	Right
Explosion3	C-3	0x3C	Sub A	Right
Exposion4	D-3	0x3E	Sub A	Right
Explosion5	E-3	0x40	Sub A	Right
Explosion6	F-3	0x41	Sub A	Right
Sm. Missile	G-3	0x43	Sub A	Right
Med. Missile	A-3	0x45	Sub A	Right
Lg. Missile	B-3	0x47	Sub A	Right
Cannon1	C-4	0x48	Sub A	Right
Cannon2	D-4	0x4A	Sub A	Right
Lg.Artillery	E-4	0x4C	Sub A	Right
M1 Fire	F-4	0x4D	Sub A	Right
Seagulls	G-4	0x4F	Sub A	Right
Crickets	A-4	0x51	Sub A	Right

Table 4: Preset 03 (MIDI Channel 03).

Sample	Note Value	Hex Value	Output Channel	Pan Setting
Rifle	C-2	0x30	Sub A	Left
Rifle Large	D-2	0x32	Sub A	Left
Rile-Auto	E-2	0x34	Sub A	Left
M-60	F-2	0x35	Sub A	Left
25mm	G-2	0x37	Sub A	Left
Explosion1	A-2	0x39	Sub A	Left
Explosion2	B-2	0x3B	Sub A	Left
Explosion3	C-3	0x3C	Sub A	Left
Exposion4	D-3	0x3E	Sub A	Left
Explosion5	E-3	0x40	Sub A	Left
Explosion6	F-3	0x41	Sub A	Left
Sm. Missile	G-3	0x43	Sub A	Left
Med. Missile	A-3	0x45	Sub A	Left
Lg. Missile	B-3	0x47	Sub A	Left
Cannon1	C-4	0x48	Sub A	Left
Cannon2	D-4	0x4A	Sub A	Left
Lg.Artillery	E-4	0x4C	Sub A	Left
M1 Fire	F-4	0x4D	Sub A	Left
Seagulls	G-4	0x4F	Sub A	Left
Crickets	A-4	0x51	Sub A	Left

Table 5: Preset 04 (MIDI Channel 04).

Sample	Note Value	Hex Value	Output Channel	Pan Setting
Rifle	C-2	0x30	SubA	Right
Rifle Large	D-2	0x32	SubA	Right
Rile-Auto	E-2	0x34	SubA	Right
M-60	F-2	0x35	SubA	Right
25mm	G-2	0x37	SubA	Right
Explosion1	A-2	0x39	SubA	Right
Explosion2	B-2	0x3B	SubA	Right
Explosion3	C-3	0x3C	SubA	Right
Exposion4	D-3	0x3E	SubA	Right
Explosion5	E-3	0x40	SubA	Right
Explosion6	F-3	0x41	SubA	Right
Sm. Missile	G-3	0x43	SubA	Right
Med.Missile	A-3	0x45	SubA	Right
Lg. Missile	B-3	0x47	SubA	Right
Cannon1	C-4	0x48	SubA	Right
Cannon2	D-4	0x4A	SubA	Right
Lg.Artillery	E-4	0x4C	SubA	Right
M1 Fire	F-4	0x4D	SubA	Right
Seagulls	G-4	0x4F	SubA	Right
Crickets	A-4	0x51	SubA	Right

Table 6: Preset 20 (MIDI Channel 05).

Sample	Note Value	Hex Value	Output Channel	Pan Setting
Rifle	C-2	0x30	SubA	Left
Rifle Large	D-2	0x32	SubA	Left
Rile-Auto	E-2	0x34	SubA	Left
M-60	F-2	0x35	SubA	Left
25mm	G-2	0x37	SubA	Left
Explosion1	A-2	0x39	SubA	Left
Explosion2	B-2	0x3B	SubA	Left
Explosion3	C-3	0x3C	SubA	Left
Exposion4	D-3	0x3E	SubA	Left
Explosion5	E-3	0x40	SubA	Left
Explosion6	F-3	0x41	SubA	Left
Sm. Missile	G-3	0x43	SubA	Left
Med. Missile	A-3	0x45	SubA	Left
Lg. Missile	B-3	0x47	SubA	Left
Cannon1	C-4	0x48	SubA	Left
Cannon2	D-4	0x4A	SubA	Left
Lg.Artillery	E-4	0x4C	SubA	Left
M1 Fire	F-4	0x4D	SubA	Left
Seagulls	G-4	0x4F	SubA	Left
Crickets	A-4	0x51	SubA	Left

Table 7: Preset 21 (MIDI Channel 06).

143

Sample	Note Value	Hex Value	Output Channel	Pan Setting
Rifle	C-2	0x30	Sub C	Right
Rifle Large	D-2	0x32	Sub C	Right
Rile-Auto	E-2	0x34	Sub C	Right
M-60	F-2	0x35	Sub C	Right
25mm	G-2	0x37	Sub C	Right
Explosion1	A-2	0x39	Sub C	Right
Explosion2	B-2	0x3B	Sub C	Right
Explosion3	C-3	0x3C	Sub C	Right
Exposion4	D-3	0x3E	Sub C	Right
Explosion5	E-3	0x40	Sub C	Right
Explosion6	F-3	0x41	Sub C	Right
Sm. Missile	G-3	0x43	Sub C	Right
Med. Missile	A-3	0x45	Sub C	Right
Lg. Missile	B-3	0x47	Sub C	Right
Cannon1	C-4	0x48	Sub C	Right
Cannon2	D-4	0x4A	Sub C	Right
Lg.Artillery	E-4	0x4C	Sub C	Right
M1 Fire	F-4	0x4D	Sub C	Right
Seagulls	G-4	0x4F	Sub C	Right
Crickets	A-4	0x51	Sub C	Right

Table 8: Preset 22 (MIDI Channel 07).

Sample	Note Value	Hex Value	Output Channel	Pan Setting
Rifle	C-2	0x30	Sub C	Left
Rifle Large	D-2	0x32	Sub C	Left
Rile-Auto	E-2	0x34	Sub C	Left
M-60	F-2	0x35	Sub C	Left
25mm	G-2	0x37	Sub C	Left
Explosion1	A-2	0x39	Sub C	Left
Explosion2	B-2	0x3B	Sub C	Left
Explosion3	C-3	0x3C	Sub C	Left
Exposion4	D-3	0x3E	Sub C	Left
Explosion5	E-3	0x40	Sub C	Left
Explosion6	F-3	0x41	Sub C	Left
Sm. Missile	G-3	0x43	Sub C	Left
Med. Missile	A-3	0x45	Sub C	Left
Lg. Missile	B-3	0x47	Sub C	Left
Cannon1	C-4	0x48	Sub C	Left
Cannon2	D-4	0x4A	Sub C	Left
Lg.Artillery	E-4	0x4C	Sub C	Left
M1 Fire	F-4	0x4D	Sub C	Left
Seagulls	G-4	0x4F	Sub C	Left
Crickets	A-4	0x51	Sub C	Left

Table 9: Preset 23 (MIDI Channel 08).

APPENDIX E: ALLEN & HEATH GL2 MIXING BOARD

This appendix serves as a guide to understanding how the Allen & Heath GL2 Route Mount Mixer is configured for use with NPSNET-3DSS. Allen & Heath products are well respected with music engineers for having perhaps the *cleanest* signal of any modestly priced mixing board. The *User Guide* for the GL2 is only somewhat helpful because of the large amount of built-in functionality. Figure 34 shows a description of the front mixing board and rear panel of the GL2. Hopefully this appendix can help to clear-up some questions concerning the use of the GL2.

A.   CONFIGURATION

The GL2 is unique among mixing boards for its use can be configured depending on its intended application. These applications include: Front-of-House, On-Stage Monitor, Recording, or Multimode which is any combination of the first three types. For use with NPSNET-3DSS, the GL2 has been configured for the Front-of-House application. The Front-of-House application allows the input of numerous audio signals, incorporates the effects produced by digital signal processors (DSPs), and allows mixing of all input signals to numerous types of outputs for use in real-time. This application is commonly used for live performances which is exactly what is needed to support the real-time constraint for adding aural cues to NPSNET. The features of the Front-of-House mode, as listed in the *GL2 User Guide*, are as follows:

- Wide range six band two sweep channel equalizer with in/out switch.
- six aux send controls with pre/post fader switching on 1-4 and 5 and 6.
- Balanced XLR Left, Right, Mono outputs.
- four balanced XLR group outputs with subgrouping to stereo.
- Comprehensive master section providing pre or post fader L-R monitoring, auto PFL/AFL, and 2-track record facility.
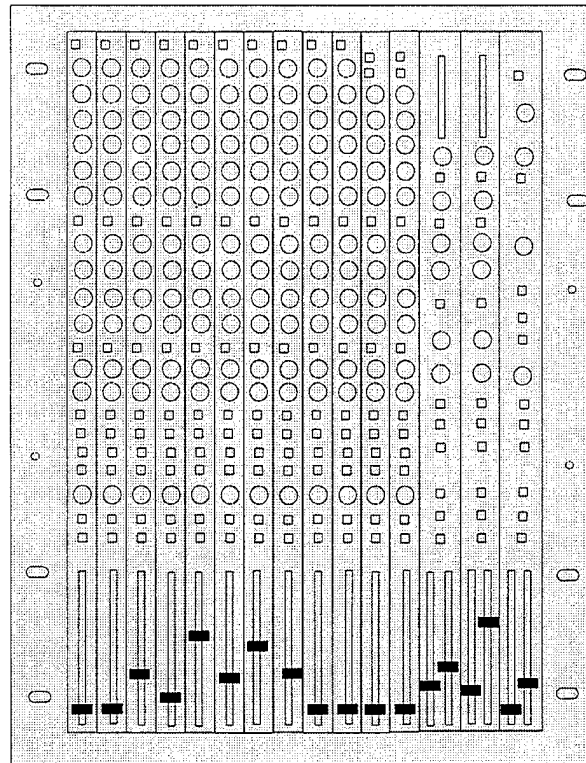
Front View of Mixing Board
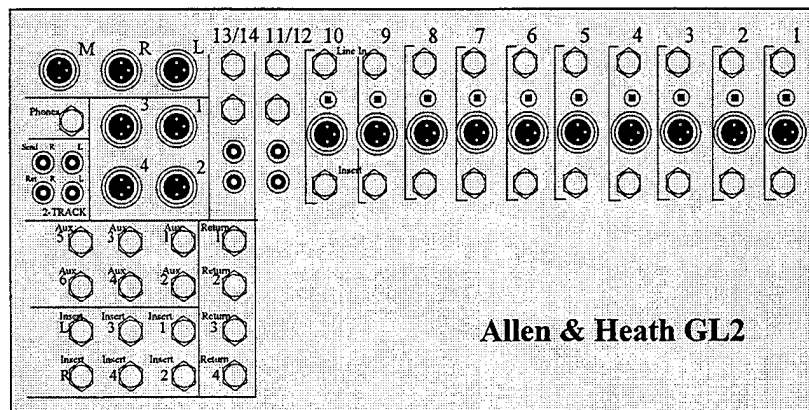
Rear Panel

Figure 34: Allen & Heath GL2 Mixing Board Front View and Rear Panel.

B.   CONNECTIONS

The GL2 supports many types of connections both input and output while in the Front-of-House mode of operation.

1. Input

The GL2 supports ten mono and two stereo input audio signals, for a total of twelve mixing channels. The mono inputs can be connected to the GL2 via *LINE* or *MIC* connectors, whereas the stereo inputs can be connected via *LINE* or *RCA* phono connectors. *LINE* refers to standard 1/4" jack (phono) and *MIC* refers to standard XLR. If *LINE* is used as opposed *MIC* (which is the case for use in NPSNET-3DSS), ensure the +48V push-button, located just above the XLR connector, are on (down position). The mono input connections are depicted in Figure 35.
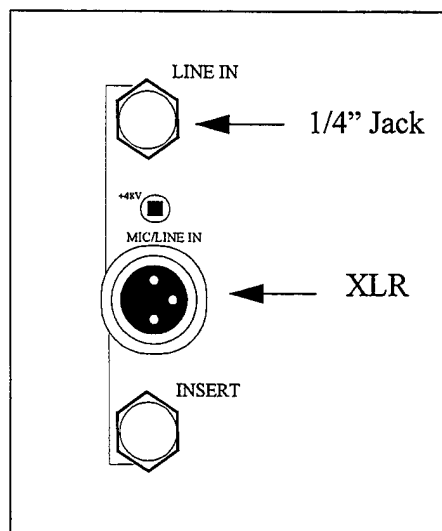
Figure 35: GL2 Mono Input Connections.

Another type of input connector is that of the insert which is also depicted in Figure 35. This connection allows an audio signal to be sent to some processing device like a DSP, in which the processed signal is returned to the same insert connection. The insert connection groups together possible send and return effects. A special cable called an insert

cable is required to take advantage of the insert connector. This cable is depicted in Figure 36.
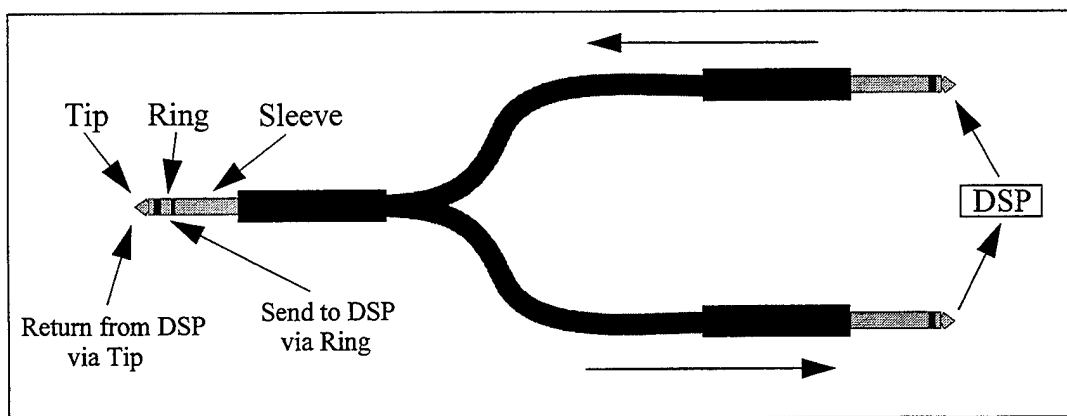
Figure 36: Insert Cable.

The last type of input is through the use of *returns*. The GL2 has four *returns* which allow new or processed signals to be mixed into the main left and right outputs. Currently, the use of *returns* is not utilized in NPSNET-3DSS.

To properly configure the GL2 for use with NPSNET-3DSS, the eight output audio signals from the EMAX II are routed to the *LINE* connectors in channels one through eight. Also, eight insert cables are connected to the insert connectors of the same channels one through eight for routing to the DP/4s.

2. Output

The GL2 supports seven *XLR*, six 1/4" jack (phono), and two RCA mono output types. As with input signals, the use of *XLR* is preferred for it is a balanced signal. One the *XLR* outputs is called *Mono*, which sums the Left and Right XLR outputs for what is called a mono mixed signal. Thus, there are actually only six *XLR* outputs which can maintain a single audio signal. The six 1/4" jack connectors are called the *sends*. These *sends* can be individually routed to amplifiers/speakers.

To properly configure the GL2 for use with NPSNET-3DSS, the *Mono* mixed output is routed to the Ramsa Subwoofer processor. The reason for this is that we do not

150

care whether the signal routed to the subwoofer processor is a left, right, or both left and right audio signal. All we are interested in is the VLF of the vehicle sounds, so the *Mono* mixed signal will suffice. The remaining six *XLR* outputs: *L, R, 1, 2, 3,* and *4* are routed to the appropriate amplifiers/speakers of the sound cube (SC). Selecting the specific output is accomplished by pressing in the proper output push-button located just above the panning knob selector on each audio channel. But at this point, we are short two outputs as required for the SC. For the last two outputs we utilize two of the six *sends*. In this case, *send 1* and *send 2* are routed to the remaining amplifiers/speakers of the SC. Each audio channel on the GL2 has six volume control knobs corresponding to the six possible *sends*. To direct output to *send 1* and *send 2*, we increase the gain on the volume control knob for *send 1* and *send 2* on the appropriate audio channels.

C.   OTHER USES

This research effort has only begun to tap the abilities of the GL2. There are no doubt better ways to maximize the effectiveness of using the GL2 for any number of possible applications. It is recommended that some sort of music engineer, such as a recording producer, give professional instruction and advice to future NPSNET sound researchers on how to configure the GL2 not only to enhance it's current application for use in NPSNET-3DSS, but also to discover other possible applications for use by the NRG.

APPENDIX F: ENSONIQ DP/4 DIGITAL SIGNAL PROCESSOR

This appendix serves as a guide to understanding how the Ensoniq DP/4 is configured for use with NPSNET-3DSS. For a more detailed understanding of how to use the Ensoniq DP/4, consult the owner's manual (see [ENSO92a] and [ENSO92b]). Calling the technical assistants from Ensoniq Corporation, the makers of the DP/4, can also be very helpful.

A.   OVERVIEW

The DP/4 is a very powerful and versatile MIDI capable effects processor. It consists of four independently programmable DSPs. The front and rear views of the DP/4 are depicted in Figure 37. It is the DP/4s which are used to produce the synthetic reverberation
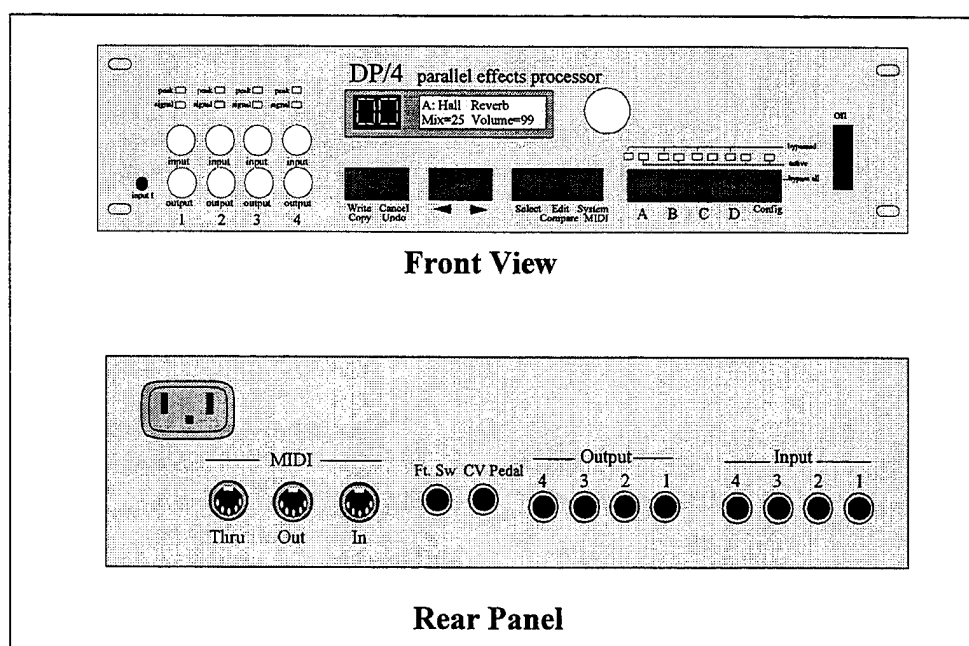
Figure 37: Ensoniq DP/4 Front View and Rear Panel.

(SR) for use in NPSNET-3DSS. The basic idea for using the DP/4s is to allocate a DSP for each of the eight audio channels required for use in the sound cube (SC). Since eight audio channels are required for the SC, we then need to use two DP/4s. The routing of the audio and MIDI signals has already be described earlier in Figure 30 on page 131. However, the

routing of the audio signals means nothing without understanding how the DP/4s are internally configured.

B.   CONFIGURATION

As depicted in Figure 38, we can see the basic overview of the DP/4. It has four
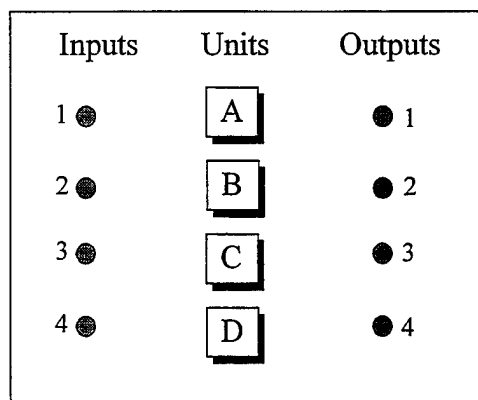
Figure 38: Basic Overview of the DP/4. After [ENSO92a].

inputs, four units (DSPs), and four outputs. Before using any functionality of the DP/4, the first step is to configure how the audio inputs are routed to the processing units (the DSPs), and how the processing units are routed to the outputs. To better understand how the DP/4 can be configured, we must think of the units (A, B, C, and D) not as a single DSP, but as both an analog-to-digital (AD) and digital-to-analog (DA) converter. As such, Figure 39 depicts some of the possible routings for which the DP/4 can be configured. The types of routings available are determined by the number of sources to be input to the DP/4. There are four possible input source configurations: one, two, three, and four source input options. An important point to remember is that the particular number of input sources selected during configuration determines the type of algorithms which can be loaded into the individual units. For use in NPSNET-3DSS, the four source configuration must be selected. After selecting the four source configuration, we now have two output choices: *Stereo Out* and *Mono Out*. In order to maintain the eight separate audio signals needed for the SC, we must select the *Mono Out* option. After selecting the four source input and the four source mono outputs in both DP/4s, we have now properly configured the DP/4s for use in
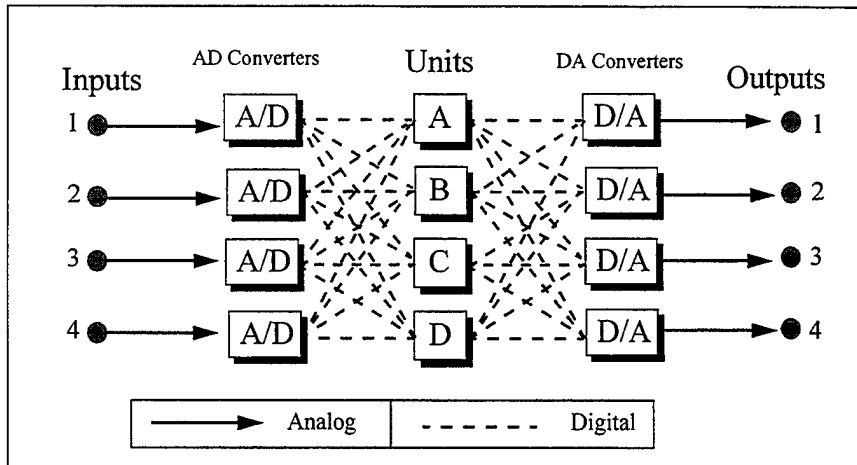
Figure 39: Possible DP/4 Configurations. After [ENSO92a].

NPSNET-3DSS. This configuration process is performed via MIDI commands during the initialization process when starting NPSNET-3DSS.

C. ALGORITHMS

Once the DP/4s have been properly configured, we need to load the appropriate algorithm into each unit (processor). The algorithm which needs to be loaded into each unit in both DP/4s is the *Large Room Rev* algorithm. This is a factory preset algorithm, but it was edited for use in NPSNET-3DSS. The *Large Room Rev* algorithm consists of thirty parameters of which the first twenty-two define the various algorithm effects, and the remaining parameters define how MIDI can access the first twenty-two parameters. The first twenty-two parameters are the same for all units, but the remaining parameters will be different in each unit depending on how MIDI will be setup to access the first twenty-two parameters. Figure 40 shows the current settings of the first twenty-two effects parameters common in each unit. The figure depicts the thirteen actual window displays of the DP4 which comprise the twenty-two effects parameters. During the initialization process when starting NPSNET-3DSS, each unit in both DP/4s is loaded with the *Large Room Rev* algorithm via MIDI commands.

155
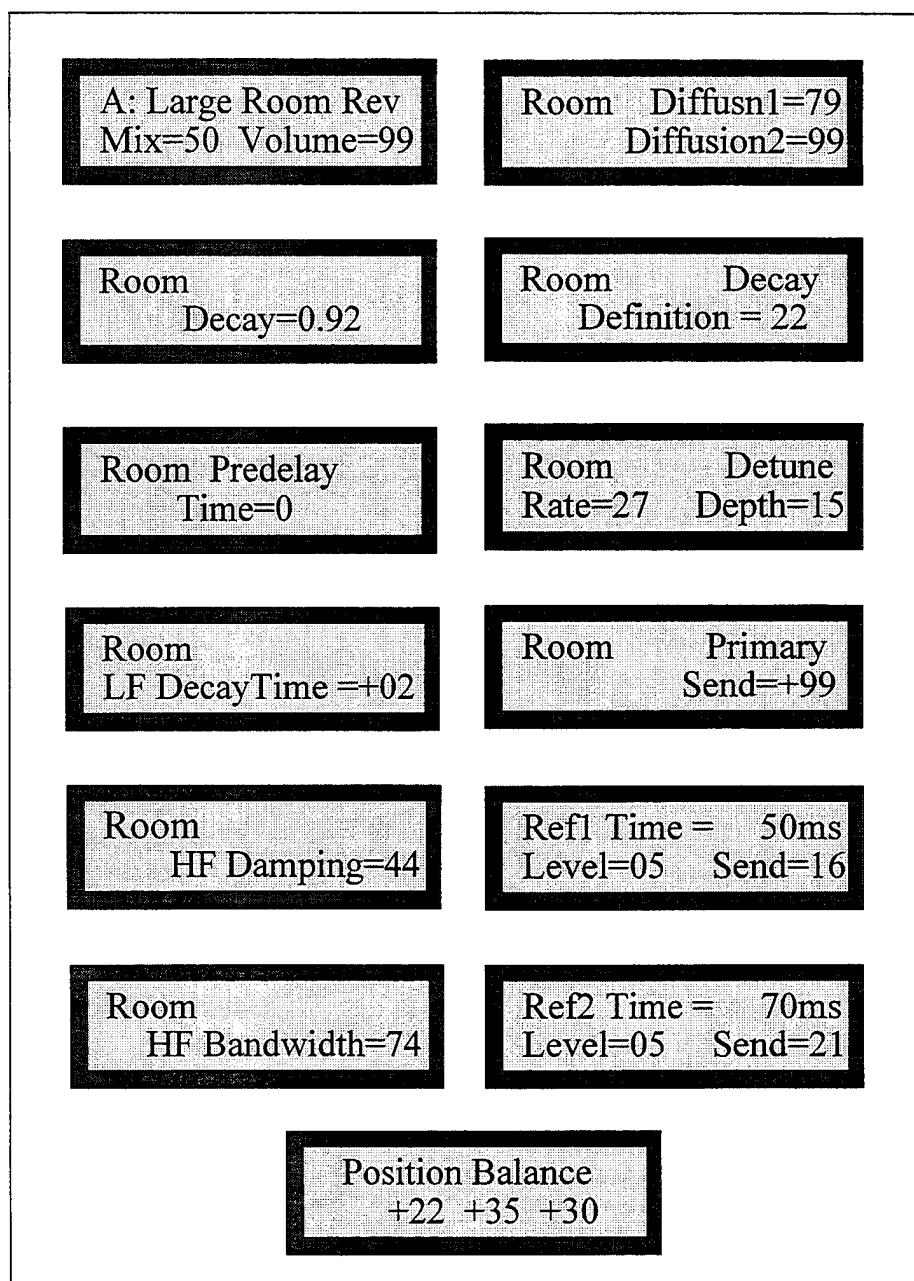
Figure 40: DP/4 Reverbertion Algorithm Effects Parameters.

D.   MIDI SETUP

There are an infinite number of ways to configure the DP/4s to respond to MIDI commands. In this research effort, the first approach taken to configure the DP/4s for MIDI

commands was to utilize MIDI System Exclusive Messages. However, this approach was unsuccessful because the DP/4s could not respond fast enough to the extra overhead bytes associated with MIDI System Exclusive Messages which were sent to the DP/4 by the faster clock speed of the Indigo. The second approach took advantage of the available sixteen MIDI channels. The basic idea of this approach is to allocate a single MIDI channel to each unit processor and control mechanism of the DP/4. As a result, a significantly smaller number of MIDI bytes need to be sent in order to control the DP/4 via MIDI. This approach was successful and the following describes this approach.

1. Unit Processors

Each of the four unit processors in both DP/4s are assigned a specific MIDI channel. As a result, if any changes need to be made to any of the units, all that is needed is a MIDI command sent on the particular unit's MIDI channel. The process of reconfiguring the DP/4s to allocate a MIDI channel to each of its four units is time consuming. However, the operating system of the DP/4 stores these changes, so the DP/4s do not need to be reconfigured prior to each use of NPSNET-3DSS. Figure 41 depicts the actual Ensoniq window displays indicating the MIDI channels selected for the individual units of DP/4 #1. Figure 42 depicts the Ensoniq window displays indicating the MIDI channels selected for the individual units of DP/4 #2.

2. Algorithms

As mentioned earlier, the last eight parameters of the *Large Room Rev* algorithm control how MIDI can access the first twenty-two effects parameters of the algorithm. The DP/4 only allows each unit to have two real-time MIDI modulation controllers assigned. Since each unit is assigned its own MIDI channel, we can assign the same MIDI modulation controllers for each algorithm loaded in the four processing units of both DP/4s. Figure 43 depicts the Ensoniq window displays indicating the particular MIDI modulation controller messages associated with their corresponding effects parameters selected for each *Large Room Rev* algorithm that is loaded in the individual units of both DP/4 #1 and #2.

157

Unit A Chan=01 MIDI is Enabled	Unit B Chan=02 MIDI is Enabled
Unit A Program Changes Received	Unit B Program Changes Received
Unit A Program Change Map Off	Unit B Program Change Map Off
Program Change 001 Selects Preset 00	Program Change 001 Selects Preset 00
Unit A Bypass= MIDI Control #075	Unit B Bypass= MIDI Control #076
Unit C Chan=03 MIDI is Enabled	Unit D Chan=04 MIDI is Enabled
Unit C Program Changes Received	Unit D Program Changes Received
Unit C Program Change Map Off	Unit D Program Change Map Off
Program Change 001 Selects Preset 00	Program Change 001 Selects Preset 00
Unit C Bypass= MIDI Control #077	Unit D Bypass= MIDI Control #078

Figure 41: DP/4 #1 Individual Unit MIDI Channel Assignments.

158

```
┌─────────────────────────────────┬─────────────────────────────────┐
│  ┌───────────────────────┐      │   ┌───────────────────────┐      │
│  │ Unit  A    Chan=07    │      │   │ Unit  B    Chan=08    │      │
│  │ MIDI  is  Enabled     │      │   │ MIDI  is  Enabled     │      │
│  └───────────────────────┘      │   └───────────────────────┘      │
│  ┌───────────────────────┐      │   ┌───────────────────────┐      │
│  │ Unit  A  Program      │      │   │ Unit  B  Program      │      │
│  │ Changes  Received     │      │   │ Changes  Received     │      │
│  └───────────────────────┘      │   └───────────────────────┘      │
│  ┌───────────────────────┐      │   ┌───────────────────────┐      │
│  │ Unit  A  Program      │      │   │ Unit  B  Program      │      │
│  │ Change  Map  Off      │      │   │ Change  Map  Off      │      │
│  └───────────────────────┘      │   └───────────────────────┘      │
│  ┌───────────────────────┐      │   ┌───────────────────────┐      │
│  │ Program Change 001    │      │   │ Program Change 001    │      │
│  │ Selects  Preset  00   │      │   │ Selects  Preset  00   │      │
│  └───────────────────────┘      │   └───────────────────────┘      │
│  ┌───────────────────────┐      │   ┌───────────────────────┐      │
│  │ Unit  A  Bypass=      │      │   │ Unit  B  Bypass=      │      │
│  │ MIDI  Control #075    │      │   │ MIDI  Control #076    │      │
│  └───────────────────────┘      │   └───────────────────────┘      │
├─────────────────────────────────┼─────────────────────────────────┤
│  ┌───────────────────────┐      │   ┌───────────────────────┐      │
│  │ Unit  C    Chan=09    │      │   │ Unit  D    Chan=10    │      │
│  │ MIDI  is  Enabled     │      │   │ MIDI  is  Enabled     │      │
│  └───────────────────────┘      │   └───────────────────────┘      │
│  ┌───────────────────────┐      │   ┌───────────────────────┐      │
│  │ Unit  C  Program      │      │   │ Unit  D  Program      │      │
│  │ Changes  Received     │      │   │ Changes  Received     │      │
│  └───────────────────────┘      │   └───────────────────────┘      │
│  ┌───────────────────────┐      │   ┌───────────────────────┐      │
│  │ Unit  C  Program      │      │   │ Unit  D  Program      │      │
│  │ Change  Map  Off      │      │   │ Change  Map  Off      │      │
│  └───────────────────────┘      │   └───────────────────────┘      │
│  ┌───────────────────────┐      │   ┌───────────────────────┐      │
│  │ Program Change 001    │      │   │ Program Change 001    │      │
│  │ Selects  Preset  00   │      │   │ Selects  Preset  00   │      │
│  └───────────────────────┘      │   └───────────────────────┘      │
│  ┌───────────────────────┐      │   ┌───────────────────────┐      │
│  │ Unit  C  Bypass=      │      │   │ Unit  D  Bypass=      │      │
│  │ MIDI  Control #077    │      │   │ MIDI  Control #078    │      │
│  └───────────────────────┘      │   └───────────────────────┘      │
└─────────────────────────────────┴─────────────────────────────────┘
```

Figure 42: DP/4 #2 Individual Unit MIDI Channel Assignments.

159

```
┌─────────────────────────────────────────────────────────────────┐
│  ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐ ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐   │
│                                                                    │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod1  Src=Cntrl-1  │       │ Mod1  Src=Cntrl-1  │           │
│    │ MIDI  Control #011 │       │ MIDI  Control #013 │           │
│    └────────────────────┘       └────────────────────┘           │
│                                                                    │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod1  Destination  │       │ Mod1  Destination  │           │
│    │  Parameter=003     │       │  Parameter=003     │           │
│    └────────────────────┘       └────────────────────┘           │
│                                                                    │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod1 Param Range   │       │ Mod1 Param Range   │           │
│    │ Min=00% Max=50%    │       │ Min=00% Max=50%    │           │
│    └────────────────────┘       └────────────────────┘           │
│                                                                    │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod2  Src=Cntrl-2  │       │ Mod2  Src=Cntrl-2  │           │
│    │ MIDI Control #012  │       │ MIDI Control #014  │           │
│    └────────────────────┘       └────────────────────┘           │
│                                                                    │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod2 Destination   │       │ Mod2 Destination   │           │
│    │  Parameter=006     │       │  Parameter=006     │           │
│    └────────────────────┘       └────────────────────┘           │
│                                                                    │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod2 Param Range   │       │ Mod2 Param Range   │           │
│    │ Min=00% Max=99%    │       │ Min=00% Max=99%    │           │
│    └────────────────────┘       └────────────────────┘           │
│            Unit A                       Unit B                     │
│  └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘ └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘   │
│  ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐ ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐   │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod1  Src=Cntrl-1  │       │ Mod1  Src=Cntrl-1  │           │
│    │ MIDI  Control #015 │       │ MIDI  Control #017 │           │
│    └────────────────────┘       └────────────────────┘           │
│                                                                    │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod1  Destination  │       │ Mod1  Destination  │           │
│    │  Parameter=003     │       │  Parameter=003     │           │
│    └────────────────────┘       └────────────────────┘           │
│                                                                    │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod1 Param Range   │       │ Mod1 Param Range   │           │
│    │ Min=00% Max=50%    │       │ Min=00% Max=50%    │           │
│    └────────────────────┘       └────────────────────┘           │
│                                                                    │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod2  Src=Cntrl-2  │       │ Mod2  Src=Cntrl-2  │           │
│    │ MIDI Control #016  │       │ MIDI Control #018  │           │
│    └────────────────────┘       └────────────────────┘           │
│                                                                    │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod2 Destination   │       │ Mod2 Destination   │           │
│    │  Parameter=006     │       │  Parameter=006     │           │
│    └────────────────────┘       └────────────────────┘           │
│                                                                    │
│    ┌────────────────────┐       ┌────────────────────┐           │
│    │ Mod2 Param Range   │       │ Mod2 Param Range   │           │
│    │ Min=00% Max=99%    │       │ Min=00% Max=99%    │           │
│    └────────────────────┘       └────────────────────┘           │
│            Unit C                       Unit D                     │
│  └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘ └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘   │
└─────────────────────────────────────────────────────────────────┘
```
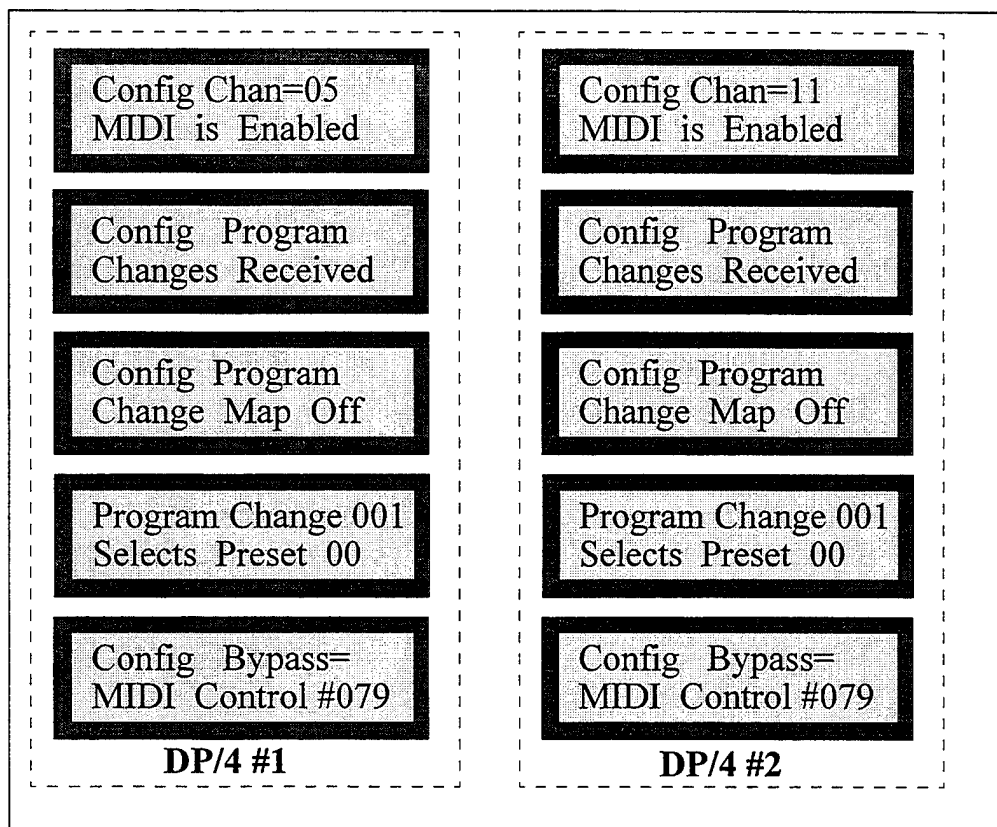
Figure 43: DP/4 #1 and #2 *Large Room Rev* Algorithm MIDI Parameter Setup.
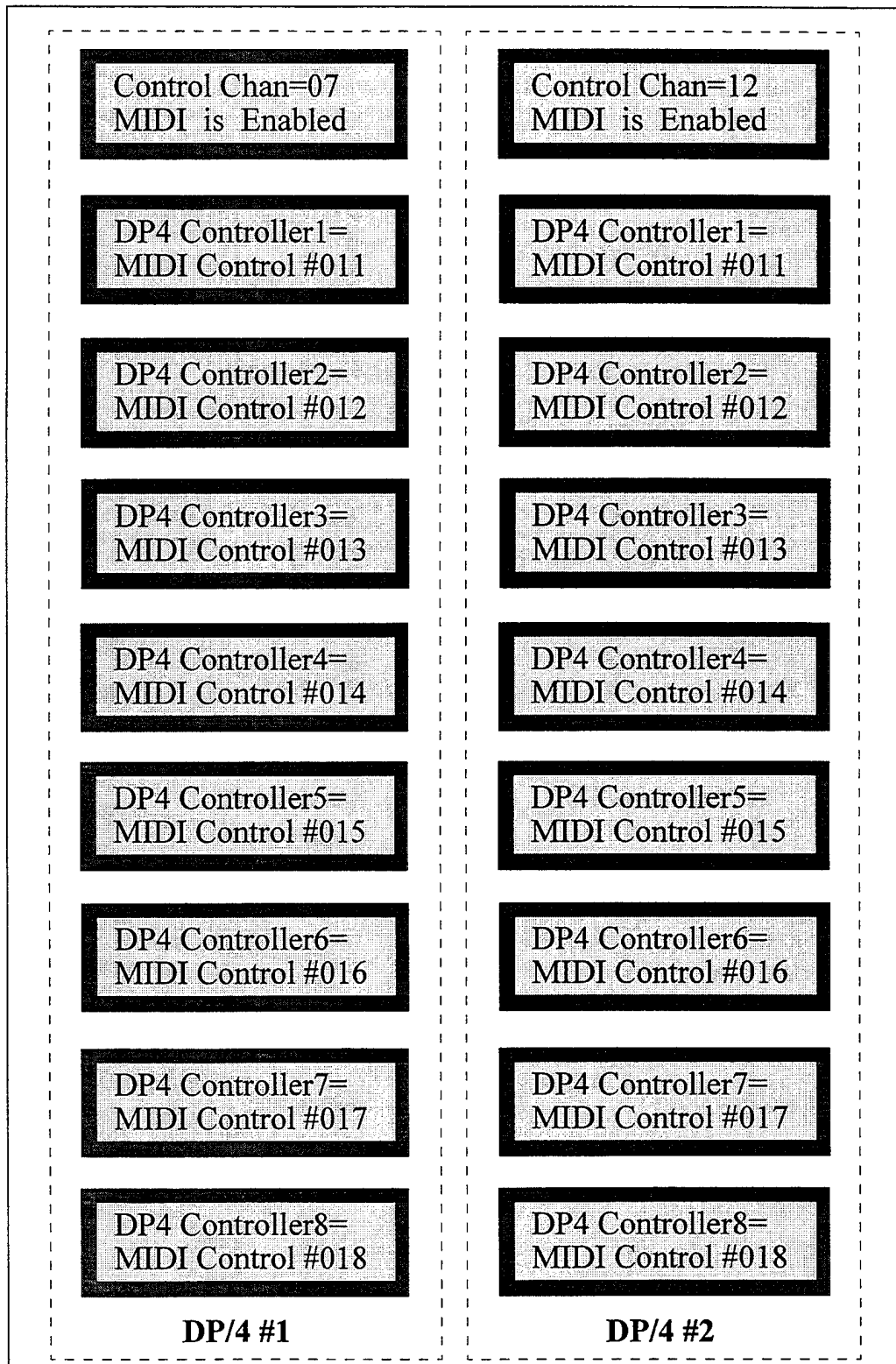
3. Configuration Channel

In order for the DP/4 to except configuration changes via MIDI, a MIDI channel is assigned to the configuration channel parameter of each DP/4. The assignments for each DP/4's MIDI configuration channel are depicted in Figure 44.

Figure 44: DP/4 #1 and #2 MIDI Configuration Channel Setup.

4. Control Channel

As in the configuration channel, in order for the DP/4 to accept control changes via MIDI, a MIDI channel is assigned to the control channel parameter of each DP/4. The assignments for each DP/4's MIDI control channel are depicted in Figure 45.

DP/4 #1	DP/4 #2
Control Chan=07 MIDI is Enabled	Control Chan=12 MIDI is Enabled
DP4 Controller1= MIDI Control #011	DP4 Controller1= MIDI Control #011
DP4 Controller2= MIDI Control #012	DP4 Controller2= MIDI Control #012
DP4 Controller3= MIDI Control #013	DP4 Controller3= MIDI Control #013
DP4 Controller4= MIDI Control #014	DP4 Controller4= MIDI Control #014
DP4 Controller5= MIDI Control #015	DP4 Controller5= MIDI Control #015
DP4 Controller6= MIDI Control #016	DP4 Controller6= MIDI Control #016
DP4 Controller7= MIDI Control #017	DP4 Controller7= MIDI Control #017
DP4 Controller8= MIDI Control #018	DP4 Controller8= MIDI Control #018
DP/4 #1	DP/4 #2

Figure 45: DP/4 #1 and #2 MIDI Control Channel Setup.

162

# APPENDIX G: BINAURAL RECORDINGS

A.   DESCRIPTION

A recording technique which captures many localization cues is that of binaural recordings. Binaural recordings are made by placing mini-microphones in a dummy head and recording some event. The recordings are then played back through headphones producing a very convincing perception of an externalized sound source as depicted in Figure 46. There are three modes of headphone listening: 1) monotic, 2) diotic, and 3)
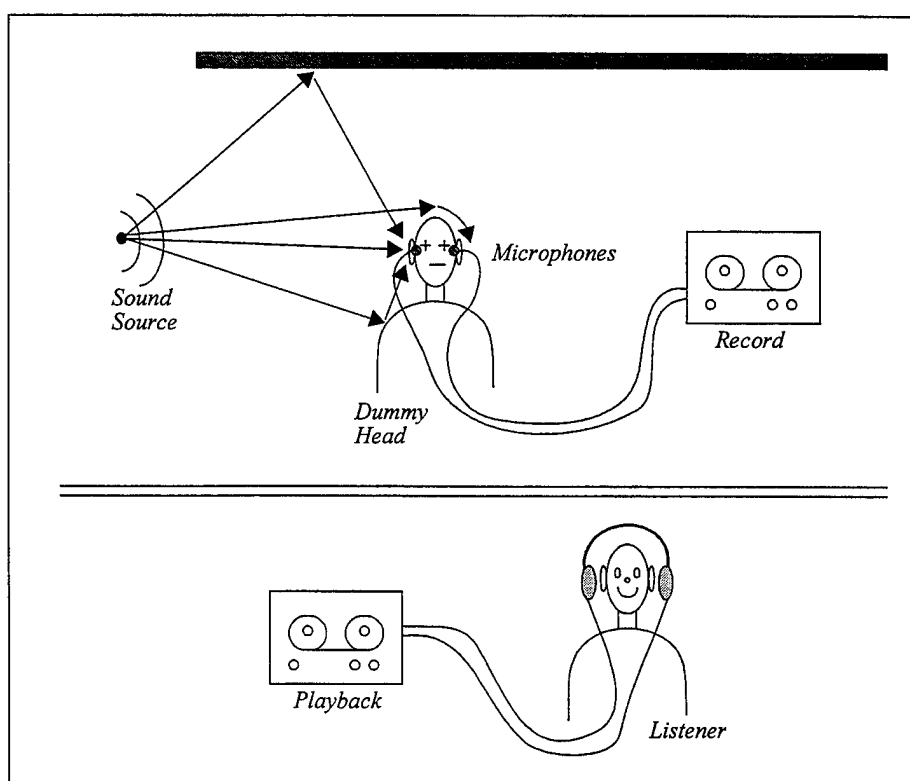
Figure 46: Binaural Recordings. From [DUDA95].

dichotic). Monotic listening refers to listening in only one ear at a time. Diotic listening refers to listening to the same sound being played in both ears. For example, listening to a mono mix recording. Dichotic listening refers to listening to different sounds being played in each ear. The monotic, diotic, and dichotic modes of headphone listening are depicted in

Figure 47. Of the three modes for headphone listening, binaural recordings are of the
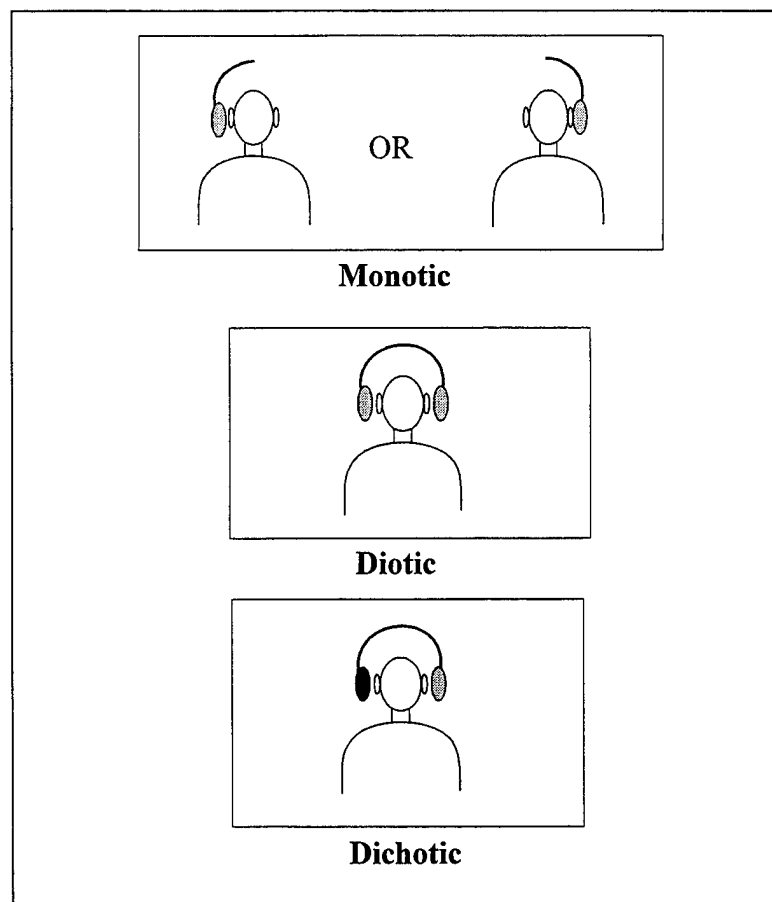
Figure 47: Modes of Headphone Listening. From [DUDA95].

dichotic mode.

B. BINAURAL RECORDING DEMONSTRATION

The following demonstration was conducted by Dr. Richard Duda from San Jose State University as part of the 1995 CCRMA Summer Workshop: *Introduction to Psychoacoustics and Psychophysics with emphasis on the audio and haptic components of virtual reality design* which was conducted at Stanford University. The students attaining the workshop (which included myself) took part in this informative binaural recording demonstration.

The instructor, Richard Duda, played a recording of a jet aircraft taking off and flying right over the top of the listener. Through headphones, he played this recording in the following formats: monaural, stereo, binaural (44 kHz sampling rate), binaural (22 kHz sampling rate), and binaural (11 kHz sampling rate). The monaural playback was totally internalized (inside the head perception) and not very spatialized. The stereo playback sounded better, but still the perception was totally internalized. The binaural (44 kHz sampling rate) playback was remarkable. The jet aircraft sounded as though it was actually flying overhead. The perception of the jet's sound was indeed externalized (outside of the head). The binaural (22 kHz sampling rate) was also externalized, but this time the elevation of the jet aircraft appeared to be lower than that of the 44 kHz sampling rate recording. The binaural (11 kHz sampling rate) was again externalized, but this time the elevation of the jet aircraft appeared to be lower than the 22 kHz sampling rate recording. Thus, it appears that the lower the sampling rate, the lower the height of the jet aircraft. This makes sense, because the lower sampling rate gives a poorer resolution of the recorded sound, and as a result, the elevation cues suffer the most. The reason that the elevation cues suffer the most is because elevation cues are much more difficult for us humans to detect than azimuth cues. Richard Duda concludes that a frequency rate above 5 kHz is needed to get elevation cues.

Another point to be made from listening to these binaural recordings is that out of the twelve people that were listening to the recordings, one person complained that he did not have any externalization of the sound of the jet aircraft. This is one of the problems of binaural recordings. A binaural recording is made from a single dummy head which is supposed to represent an average sized human head. The problem with this is that not everyone has an averaged size head. So, it is important to remember that binaural recordings are not guaranteed to work for everyone. Furthermore, when listening to binaural recordings, Richard Duda recommends using closed-end headphones.

C. BINAURAL RECORDING CD'S

A place to obtain binaural recordings on CDs is the following:

The Binaural Source
Recordings for Headphone Experiences
BOX 1727
Ross, CA 94957
(800) 934-0442

APPENDIX H: SOUND PERCEPTION EXPERIMENTS

This appendix contains information on various sound localization and echo experiments principally conducted by Brent Gillespie as part of the instruction during the 1995 CCRMA Summer Workshop: *Introduction to Psychoacoustics and Psychophysics with emphasis on the audio and haptic components of virtual reality design* at Stanford University. The students attending the workshop (which included myself) were the test subjects.

To localize sound, we humans use three main sound cues: 1) intensity, 2) delay, and 3) reverberation [GILL95c]. The following experiments help to reveal how we use these cues to localize sound.
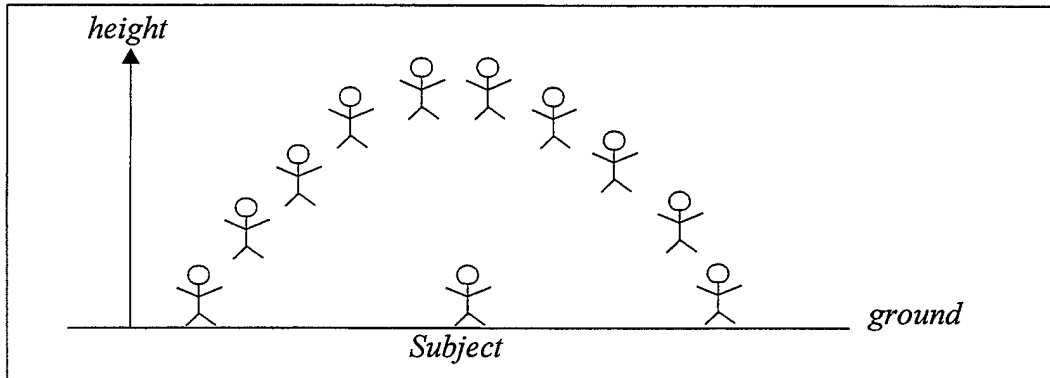
A.   LATERAL LOCALIZATION EXPERIMENT

A person (the subject) sat in the middle of a large room with his eyes closed. Five people were then spaced evenly apart in a straight-line in front of the subject, and five people were placed evenly apart in a straight-line in back of the subject. The various people in the line then shook their car keys at random, and the subject was asked to point in the direction of the sound. The experiment was repeated with the subject folding his ears flat against his head. The experiment showed that the subject could better distinguish/localize sounds with the normal use of his ears, as opposed to folding over his ears.
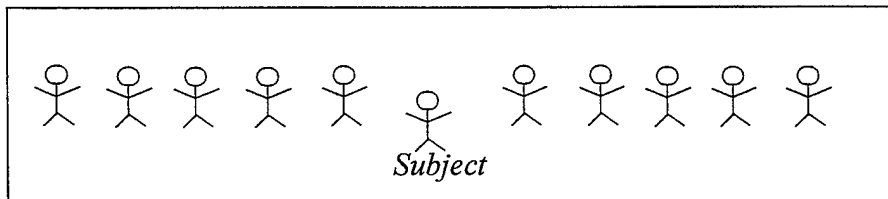
Figure 48: Lateral Localization Experiment.

B. VERTICAL LOCALIZATION EXPERIMENT

The same person (the subject) sat in the middle of the same large room with his eyes closed. Ten people were then evenly spaced in a semicircle placed vertically over the subject's head. The ten people in the semicircle then randomly shook their car keys again. Again, the subject was asked to point in the direction of the sound. The experiment showed that the subject was not as accurate in locating the correct direction of the sound in the vertical plane. The experiment was again repeated with the subject folding his ears flat against his head. This time the subject had great difficulty in correctly localizing the proper direction of the sound source.


Figure 49: Vertical Localization Experiment.

C. LATERAL DISTANCE PERCEPTION EXPERIMENT

One person (the subject) sat with her eyes closed at one end of the room. Ten people were then spaced evenly apart in a straight-line extending outward from the subject. The ten people were numbered from 1 to 10 with 1 being closest to the subject. The ten people then randomly shook their car keys again. The subject was asked to state the number of the person making the noise. The experiment showed that the subject could distinguish distances, but only for large resolutions. The subject could not distinguish small resolutions

between each person. For example, the subject could distinguish a sound as coming from somewhere in the range between person 3 and 6, but not exactly at say 3 or 4.
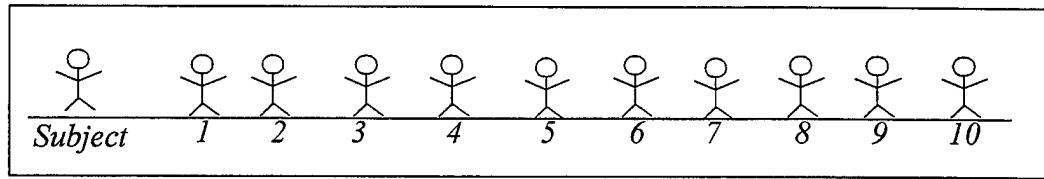

Figure 50: Distance Perception Experiment.

D.   ECHO EXPERIMENT

This experiment was conducted in two parts: (1) outdoors and (2) indoors. In part (1), a group of people was placed outside at an arbitrary distance from the wall of a tall building. Next, another individual, located with this group of people, slapped together two large pieces of aluminum. As a result, a loud metallic-like clap sound was heard by the group followed by its echo off the wall of the building. This individual then walked a few paces toward the building (away from the group of people) and again produced the loud clap sound. Again, there was an echo heard by the group of people. This procedure was continued until the group could not longer perceive any echo. The spot where the clap produced no perceivable echo was measured off in paces from the wall of the building. This distance was 6 paces. Next, the distance from the wall to the group of people was also measured in paces. This distance was 38 paces. The sound of the clap heard by the group of people had two distances to travel. One is the direct route from individual to the group of people. The other is the indirect route from the individual to the wall and then reflected off the wall back to the group of people. Using these distances, we found that the sound which traveled the further distance was delayed by approximately 34 milliseconds. Thus, 34 ms is the threshold at which we begin to perceive an echo in this outdoors experiment.

Part (2) of this experiment was conducted inside a large room. However, in this part of the experiment, a computer was used to simulate a clap sound followed by its echo. The computer gradually shortened the length between the clap and its echo. The same group

169

was then asked to determine when there was no perceptual echo. Under these inside conditions, the threshold at which the group could perceive an echo was 5 milliseconds.

Although there is lots of room for error in this experiment, there is still a significant difference between our perception of echo outdoors as opposed to indoors. This suggests many things, but one possibility is that our ability to localize sound might also be based on preconceived notions of what we think we should be hearing in different ambient conditions.
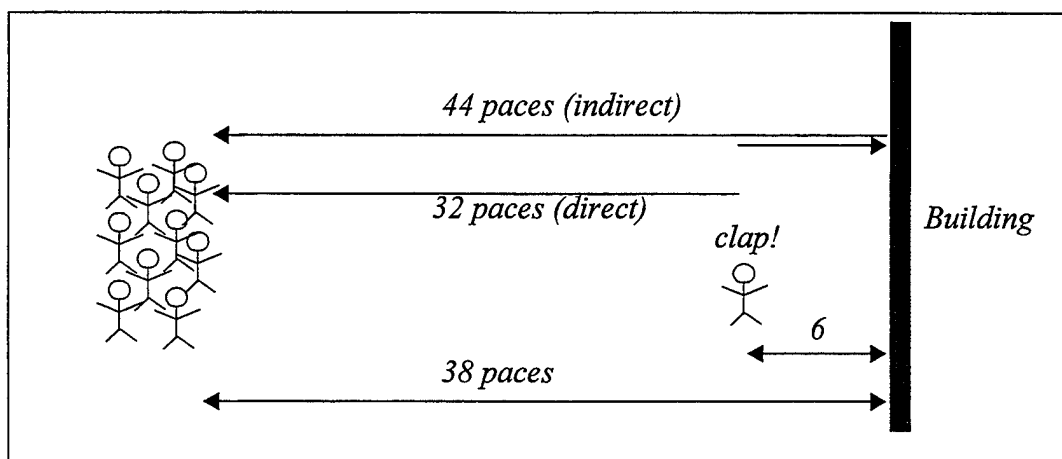
Figure 51: Echo Experiment (outdoors).

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center..2
    Cameron Station
    Alexandria, VA 22304-6145

2. Dudley Knox Library, Code 013 2
    Naval Postgraduate School
    Monterey, CA 93943-5002

3. LTC James Wall 5
    Army Research Laboratory
    Aberdeen Proving Ground, MD 21005

4. Chairman, Code 32 CS 2
    Computer Science Department
    Naval Postgraduate School
    Monterey, CA 93943-5000

5. Dr. Michael J. Zyda, Code CS/Zk 10
    Computer Science Department
    Naval Postgraduate School
    Monterey, CA 93943-5000

6. Dr. David R. Pratt, Code CS/Pr...2
    Computer Science Department
    Naval Postgraduate School
    Monterey, CA 93943-5000

7. Dr. Don Brutzman, Code UW/Br 2
    Computer Science Department
    Naval Postgraduate School
    Monterey, CA 93943-5000

8. John S. Falby, Code CS/Fa...2
    Computer Science Department
    Naval Postgraduate School
    Monterey, CA 93943-5000

9. Paul T. Barham, Code CS..1
    Computer Science Department
    Naval Postgraduate School
    Monterey, CA 93943-5000

10. Dr. Elizabeth M. Wenzel 1
    NASA-Ames Research Center
    Moffet Field, CA 94035-1000